The Journal of Network Theory in Finance EDITORIAL BOARD

Editor-in-Chief

KIMMO SORAMÄKI Financial Network Analytics Ltd.

Associate Editors

FRANKLIN ALLEN Brevan Howard Centre & Imperial College Business School IGNAZIO ANGELONI European Central Bank STEFANO BATTISTON University of Zurich ANDREW COBURN RMS, Inc. & University of Cambridge RAMA CONT Imperial College London & CNRS (France) BEN CRAIG Deutsche Bundesbank RODNEY GARRATT Federal Reserve Bank of New York **CO-PIERRE GEORG** Deutsche Bundesbank ANDREW G. HALDANE Bank of England SERGEY IVLIEV Prognoz DROR KENETT Boston University IMAN VAN LELYVELDT De Nederlandsche Bank THOMAS LUX University of Kiel & University Jaume I TIZIANA DI MATTEO King's College London ROSARIO NUNZIO MANTEGNA Central European University & Palmero University PETER SARLIN Hanken School of Economics DIDIER SORNETTE ETH Zurich MURAT UNAL SONEAN GmbH & Funds at Work

SUBSCRIPTIONS

The Journal of Network Theory in Finance (Print: ISSN 2055-7795 and Online: ISSN 2055-7809) is published four times per year by Incisive Media, Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK. Subscriptions are available on an annual basis, and the rates are set out in the table below.

	UK	Europe	US
Risk.net Journals	£1850	€2200	\$2950
Print only	£699	€979	\$1155
Risk.net Premium	£2750	€3300	\$4400

Academic discounts are available. Please enquire by using one of the contact methods below.

All prices include postage. All subscription orders, single/back issues orders, and changes of address should be sent to:

UK & Europe Office: Incisive Media (c/o CDS Global), Tower House, Sovereign Park, Market Harborough, Leicestershire, LE16 9EF, UK. Tel: 0870 787 6822 (UK), +44 (0)1858 438 421 (ROW); fax: +44 (0)1858 434958

US & Canada Office: Incisive Media, 120 Broadway, 5th Floor, New York, NY 10271, USA. Tel: +1 212 457 9400; fax: +1 646 417 7705

Asia & Pacific Office: Incisive Media, 20th Floor, Admiralty Centre, Tower 2, 18 Harcourt Road, Admiralty, Hong Kong. Tel: +852 3411 4888; fax: +852 3411 4811 Website: www.risk.net/journal E-mail: incisivehv@subscription.co.uk

The Journal of Network Theory in Finance GENERAL SUBMISSION GUIDELINES

Manuscripts and research papers submitted for consideration must be original work that is not simultaneously under review for publication in another journal or other publication outlets. All articles submitted for consideration should follow strict academic standards in both theoretical content and empirical results. Articles should be of interest to a broad audience of sophisticated practitioners and academics.

Submitted papers should follow *Webster's New Collegiate Dictionary* for spelling, and *The Chicago Manual of Style* for punctuation and other points of style, apart from a few minor exceptions that can be found at www.risk.net/journal. Papers should be submitted electronically via email to: journals@incisivemedia.com. Please clearly indicate which journal you are submitting to.

You must submit two versions of your paper; a single LATEX version and a PDF file. LATEX files need to have an explicitly coded bibliography included. All files must be clearly named and saved by author name and date of submission. All figures and tables must be included in the main PDF document and also submitted as separate editable files and be clearly numbered.

All papers should include a title page as a separate document, and the full names, affiliations and email addresses of all authors should be included. A concise and factual abstract of between 150 and 200 words is required and it should be included in the main document. Five or six keywords should be included after the abstract. Submitted papers must also include an Acknowledgements section and a Declaration of Interest section. Authors should declare any funding for the article or conflicts of interest. Citations in the text must be written as (John 1999; Paul 2003; Peter and Paul 2000) or (John *et al* 1993; Peter 2000).

The number of figures and tables included in a paper should be kept to a minimum. Figures and tables must be included in the main PDF document and also submitted as separate individual editable files. Figures will appear in color online, but will be printed in black and white. Footnotes should be used sparingly. If footnotes are required then these should be included at the end of the page and should be no more than two sentences. Appendixes will be published online as supplementary material.

Before submitting a paper, authors should consult the full author guidelines at:

http://www.risk.net/static/risk-journals-submission-guidelines

Queries may also be sent to:

The Journal of Network Theory in Finance, Incisive Media, Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK Tel: +44 (0)20 7004 7531; Fax: +44 (0)20 7484 9758 E-mail: journals@incisivemedia.com

The Journal of

Network Theory in Finance

The journal

The Journal of Network Theory in Finance is an international refereed journal focusing on the application of network theory in finance.

Financial institutions and markets are highly interconnected but only recently has a burgeoning literature started to emerge to map these interconnections and to assess their impact on financial risks and returns. *The Journal of Network Theory in Finance* is an interdisciplinary journal publishing academically rigorous and practitionerfocused research. It brings together studies carried out in often disconnected areas within academia and other research institutions, by policymakers and by industry practitioners.

The Journal of Network Theory in Finance has three fundamental aims:

- (1) to foster high-quality, original and innovative work;
- (2) to provide practitioners, policymakers and academics with access to the resulting research; and
- (3) to serve as an educational forum on timely issues concerning network theory in finance.

Content Guidelines

The journal publishes data-driven or theoretical work that can be in, but is not limited to, the following areas:

- flows in financial networks (payment, trade, securities),
- financial exposure of networks (interbank, sectoral),
- network models of systemic risk,
- network-based models in finance and insurance,

- interdependencies in financial markets,
- network-based models for asset allocation and portfolio management,
- asset correlation networks,
- social network analysis (SNA) in finance,
- network visualization in finance,
- using network analysis for extracting meaningful insights from large-scale financial data.

Research Papers

Research papers are typically full-length academic papers of between 8,000 and 10,000 words in length. Topics for research can be of interest to both academic and industry professionals. Papers submitted for publication must be original and should not have been published or considered for publication elsewhere.

CONTENTS

Letter from the Editor-in-Chief	vii
RESEARCH PAPERS	
Eccentricity in asset management Hakan Kaya	1
Emergence of the EU corporate lending network Grzegorz Hałaj, Urszula Kochańska and Christoffer Kok	33
Risk diversification: a study of persistence with a filtered correlation-network approach Nicoló Musmeci, Tomaso Aste and T. Di Matteo	77
A multiplex network analysis of the Mexican banking system: link persistence, overlap and waiting times José-Luis Molina-Borboa, Serafin Martínez-Jaramillo, Fabrizio López-Gallo and Marco van der Leij	99

Editor-in-Chief: Kimmo Soramäki Publisher: Nick Carver Journals Manager: Dawn Hunter Editorial Assistant: Carolyn Moclair Marketing Executive: Giulia Modeo Subscription Sales Manager: Iliana Damoutzidou Global Head of Sales: Michael Lloyd Head of Product & Delegate Sales: Michelle Godwin Composition and copyediting: T&T Productions Ltd Printed in UK by Printondemand-Worldwide

©Incisive Risk Information (IP) Limited, 2015. All rights reserved. No parts of this publication may be reproduced, stored in or introduced into any retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the copyright owners.



LETTER FROM THE EDITOR-IN-CHIEF

Kimmo Soramäki

Financial Network Analytics Ltd.

The Journal of Network Theory in Finance is being launched at a time of everincreasing connectivity and complexity in financial markets and in the financial system. This complexity and connectivity is currently being tackled by regulators, who collect ever-more-detailed information on links between financial institutions for macroprudential supervision, by asset managers, who need to understand the complex time-varying dynamics of interdependencies in financial markets, and by risk managers, who want to understand how emerging systemic risks can be identified as they cascade through the financial markets.

In recent years, network theory has proven useful in applications ranging from cancer research to the social graph. Applications of network theory are quickly becoming ever more present in finance, with network analysis providing answers to questions where traditional analysis methods are weak, and leading to improved models across wide types of financial risks. In fact, networks underlie virtually every type of risk, including liquidity, operational, insurance and credit risk.

The Journal of Network Theory in Finance aims to bring together research currently carried out in disparate areas at universities and by policymakers and industry practitioners. This research has often been published in a wide variety of journals across physics, finance, economics and other disciplines, or it remains unpublished due to the avant garde nature of the field. The publisher and the editorial board therefore see great value in launching this interdisciplinary journal for publishing academically rigorous and practitioner-focused research on the application of network theory in finance and related fields.

On September 23, 2014 a conference entitled "Network Theory and Financial Risk" was held at the Centre for Risk Studies at Cambridge University to inaugurate the journal, and submissions to that conference form the backbone of the journal's first issue. The papers reflect the scope of the journal well, ranging from applications in asset management to the measuring of counterparty exposures between financial institutions.

The first paper in the issue, "Eccentricity in asset management" by Hakan Kaya, considers connectivity among financial assets and investigates whether a node's position in the network can predict the magnitude of the asset's returns, and whether the network structure can explain systemic events. The author finds that assets that are located near the center of the network tend to have higher returns and shows that an investment strategy based on this information has historically provided value.

Importantly, the paper shows how methodologies from network theory can help reformulate long-standing questions in finance and provide new insights.

Understanding interbank exposures has been a focus of financial stability analysis in recent years. The issue's second paper, "Emergence of the EU Corporate Lending Network" by Grzegorz Hałaj, Urszula Kochańska and Christoffer Kok, extends the analysis to bank–firm relationships by developing a network formation algorithm that estimates these relationships based on largely public data. The authors find that contagion can be amplified when this transmission channel is taken into account. The work will likely find direct applications in stress testing both by regulators and at banks.

Our third paper, "Risk diversification: a study of persistence with a filtered correlation-network approach" by Nicoló Musmeci, Tomaso Aste and T. Di Matteo, addresses the problem of correlation structures observed in the past not always persisting into the future. Finding persistent structures is important for risk diversification and the authors develop a new clustering algorithm to identify such structures. Clusters of assets identified from correlation structures are more economically meaningful and can be more stable than, for example, industry-based classifications, and they are therefore important for better investment and risk management decisions.

Much of the initial analysis on financial interlinkages has focused on particular financial instruments for which data has been available. The issue's fourth paper, "A multiplex network analysis of the Mexican banking system: link persistence, overlap and waiting times" by José-Luis Molina-Borboa, Serafin Martínez-Jaramillo, Fabrizio López-Gallo and Marco van der Leij, overcomes this simplification and specifically studies the interaction between several layers of interconnectivity across markets (repos, uncollateralized loans, cross holdings, etc). The paper provides very rich insight into the complex multiplex nature of the Mexican financial system and will help other researchers understand and model how these networks interact in other countries where data at such a detailed level is not available, or where the data is confidential.

Last but not least, on behalf of the editorial board I welcome readers to the inaugural issue of *The Journal of Network Theory in Finance*. I hope the journal will foster both weak and strong links among the community of researchers interested in financial networks, and that the techniques and findings presented therein find valuable application among practitioners.



Research Paper Eccentricity in asset management

Hakan Kaya

Neuberger Berman, 605 Third Avenue, New York, NY 10158, USA; email: hakan.kaya@nb.com

(Received August 21, 2014; revised December 5, 2014; accepted December 17, 2014)

ABSTRACT

We describe how networks based on information theory can help measure and visualize systemic risk, enhance diversification, and help price assets. To do this, we first define a distance measure based on the mutual information between asset pairs and use this measure in the construction of minimum spanning trees. The dynamics of the shape and the descriptive statistics of these trees are analyzed in various investment domains. The method provides evidence of regime changes in dependency structures prior to market sell-offs and, as such, it is a potential candidate for monitoring systemic risk. We also provide empirical evidence that the assets that are located toward the center of the network tend to have higher returns. Finally, an investment strategy that utilizes network centrality information is shown to add value historically.

Keywords: risk measurement; minimum spanning tree; systemic risk; entropy; mutual information; eccentricity.

1 MOTIVATION

The worst is behind us.

Richard Fuld (final Chairman and CEO of Lehman Brothers)

Little did he know how interconnected the finance world had become and what this connectedness meant for his company. Richard Fuld gave the above quote (see Fuld

Print ISSN 2055-7795 | Online ISSN 2055-7809 Copyright © 2015 Incisive Risk Information (IP) Limited

2008) to the news media in the wake of the successive write-offs in Wall Street investment banks in 2008. Since then, ever-increasing synchronization between assets, asset classes and strategies has been one of the central themes for the post-2008 asset management industry. This theme is now a part of our lives as it is a ubiquitous feature in almost all domains of risk. It is in market returns, between corn and crude oil, in default probabilities, in central bank policies, and therefore it is now in the business cycles of allegedly independent countries.

Now that we have to live in this so-called new normal, in which tail risks not only happen more frequently than before but also happen in many aspects of the investment world at the same time, perhaps expectedly, we have seen an increasing amount of research effort focusing on the domestication of synchronization. Yet it is puzzling to see that when it comes to measuring this key concept of risk, we still do not have a clear-cut, valid and broadly accepted definition.

A study of synchronization often includes three stages. First, we define a pairwise similarity measure between underlying constituents. Then a structural model is assumed in order to summarize the potentially large amount of information in an efficient and perhaps visual manner. Finally, time series properties of the outputs of this model are used to test hypotheses, hopefully to show that synchronization defined and modeled in the study can predict systemic events and/or warn of trouble ahead. The outcome is an online monitor and a set of rules of thumb that policy makers, asset managers and other interested persons can use to track the probability of contagious and catastrophic events for enhanced risk taking.

On the definition of synchronization, financial literature provides us with a number of perspectives. An extensive up-to-date review can be found in Bisias *et al* (2012). The majority of these studies nest in the traditional setting where randomness is represented by bell-shaped (normal) distributions and similarities are modeled through correlation coefficients measuring merely linear dependencies. For example, Engle and Kelly (2012) takes the mean of all cross correlations, and Kritzman *et al* (2010) work on the eigenvalues of a correlation matrix. The contribution to risk type frameworks detailed in studies such as Adrian and Brunnermeier (2011) and Acharya *et al* (2010) also build on linear relationships. Deviating from linearity, Diebold and Yilmaz (2009) introduced variance decompositions for measuring financial connectedness of firms, however the dependence of their systemic index on the ordering of the variables they used in vector autoregressions drew criticism (see Klößner and Wagner 2014).

The quick and dirty way, which is widespread in media and sell-side research (see Kolanovic *et al* 2011) is imposing a summarizing structure by taking the average of all measures of pairwise relationships. Although these studies are certainly of interest, their value in the financial-market contexts is limited to stressing correlations converging to unity and/or regression betas (for CAPM or related factor models) becoming more important than before.

A richer context has started to emerge with Mantegna (1999) introducing networkbased hierarchical clustering in financial markets. While still mostly based on correlation distances, these networks allowed a more intuitive understanding of the synchronization, with visual aids making it possible to question the plausibility and robustness of otherwise complex interrelationships. This line of research has since flourished so much that we cannot list all of the important references. Nevertheless, interested readers can consult Hiemstra and Jones (1994), Boginski *et al* (2005), Kenett *et al* (2010), Billio *et al* (2012), Kenett *et al* (2012), Lautier and Raynaud (2013), Sandoval (2014), Barigozzi and Brownlees (2013), and Majdandzic *et al* (2014) for early as well as recent treatments and further examples.

The normative implications of these new tools have concentrated more on systemic risk measurement as we listed above, and less on asset pricing and portfolio allocation. Recently, Ozsoylev and Walden (2011) and Buraschi and Porchia (2012) have studied the asset pricing aspect and the latter concluded that "central" firms have lower P/D ratios and higher expected returns. A rare and tangible portfolio allocation rule has been offered and backtested in Kritzman *et al* (2010), which recommended shifting toward bonds from equities during stress that is screened through an increasing trend in their systemic risk measure.

In a complementary way to the aforementioned research, in this paper our overarching goal will be to examine the dynamics of the synchronization between assets, and further the understanding of what this means for their risks, returns and for asset allocation in general. We aim to achieve this goal step by step.

First, we will introduce the mutual information distance as our dissimilarity measure. This information theory based concept is similar to the correlation coefficient, but it does not suffer from parameterization and linearity. In its essence, estimating mutual information is about measuring how independent a pair of time series are and in most cases assumptions about the underlying distributions of time series are not needed. In other words, mutual information is model free.

Second, this distance measure will be used to represent the pairwise similarity information in a network framework. In studies with a large number of assets, these networks themselves are not that helpful. What is needed is a robust and fast enough information compression tool that can summarize the essence of the information in networks in a unique way. For this purpose, we will utilize minimum spanning trees (MSTs), which are subsets of the initially fully connected mutual information networks. As their name suggests, MSTs span the asset universe connecting every asset to every other asset without cycles and with minimum total network distance between assets. To draw an analogy, we can imagine a hypothetical country trying to decide how to optimally build roads (edges of a network) between all of its cities (nodes of a network) with no cycles and with minimum cost of construction. In this case, the governors of this country need to solve an MST problem. Note at this point

that MSTs are neither the only nor the most efficient network compression tools. A wide variety of methods exist. See, for example, Tumminello *et al* (2010) for a set of related techniques. We chose to work with MSTs due to their intuitive construction, robustness and speed for handling large number of assets.

Third, having a network opens up many location-related questions. For example, we need a coordinate system to describe the whereabouts of a node in the network. This addressing issue requires the determination of a reference point; a center. There are no easy and optimal answers for these choices and it is far beyond the scope of this paper to find out what is the right answer. Instead, to describe the centrality of assets we will rely on an easy to grasp concept called eccentricity (Hage and Harary 1995). Eccentricity of a node is simply the shortest path from the farthest out node. It is calculated by measuring the distances from the originating node to all other nodes and taking the maximum. The lower the eccentricity of a particular node, the closer it is to every other node, that is, the more central it is. A node with the lowest eccentricity score will be called the center of the network, and the eccentricity of this central node will be given a special name: radius. Again, eccentricity is neither the only nor the most important network centrality measure. Many others exist, and interested readers can consult Koschützki and Schreiber (2004, 2008).

Finally, we will put these concepts into use and analyze the eccentricities of nodes in mutual information networks through time. We will define our own systemic risk index called the average eccentricity and test its efficacy in warning of crashes. Next, we will carry out simple asset pricing tests to show whether eccentricity centrality can cross-sectionally discriminate returns. We will provide this analysis in two domains: first, in a global asset allocation setting, and second, in a sector/region equity allocation setting. Backtests will be provided to assess the value added by the eccentricity based information in both investment universes.

2 METHODOLOGY

The construction of a mutual-information-minimum-spanning-tree-eccentricitycentrality measure requires us to define a set of objects. Although these concepts can be found in elementary statistics, information theory and graph theory books, for completeness we will provide the necessary descriptions in what follows. We will start with mutual information, and move on to the definition of MSTs. Finally, we will define the eccentricity of a node and the average network eccentricity of a fully connected graph.

2.1 Mutual information distance

Nonlinearity in financial returns has been extensively studied and empirically wellsupported. See, for example, Rothman (1999) for the treatment of various sources of



FIGURE 1 Example of a nonlinear relationship with zero correlation.

An example in which linear or rank correlation measures fail to capture the underlying nonlinear relationship. Pearson (linear) correlation (x, y): 0.00. Kendall (rank) correlation (x, y): 0.00. Mutual information (x, y): 0.91.

nonlinearity in returns from time-changing variance to asymmetric cycles, and from higher-moment structures, to thresholds, breaks and regime switches.

Given these issues, what motivates us to use mutual information, an informationtheoretic construct, as a measure of dependence is its ability to capture both linear and nonlinear dependencies without requiring any assumption of a specific model. To demonstrate what this means, let us consider the simple dependence structure exhibited in Figure 1. The relationship, which is created by adding noise to a cosine wave, exhibits a clear nonlinear dependence. While this is true both by construction and by visual inspection, the sample estimates of both the linear and the rank correlation coefficients are zero implying a lack of relationship. Needless to say, had we assumed a bivariate normal distribution, we would have mistakenly concluded that the underlying random variables were indeed independent. In contrast, the mutual information between these random variables is a relatively large nonzero number implying a strong relationship is present.

Additionally, as opposed to the linear coefficient of correlation, the calculation of mutual information does not require the computation of means, variances and covariances. This is a key advantage as the computation of these moments is often challenging when the underlying data is composed of financial time series. This and

www.risk.net/journal

various other desirable properties are also stressed in several studies which can be found in Dionisio *et al* (2004). As a result, one would expect mutual information to be more robust compared with the traditional measures of dependence.

As well as these advantages, mutual information has some drawbacks. Most importantly, to calculate this measure, we need information on the joint and marginal probability density functions of the underlying random variables. The approximation of such distributions by histogram methods can create bias problems as studied in Moon *et al* (1995), Kraskov *et al* (2004) and in Walters-Williams and Li (2009). Therefore, it is important that we use an efficient estimator when studying dependence with mutual information and carry out robustness checks with different estimation methods.

Cover and Thomas (2006) covers the basic definitions, axioms, and properties required for the development of information theory which is built on the concept of the entropy of a random variable. They describe entropy of a random variable as the measure of the amount of information required on the average to describe the random variable itself. On the one hand, a random variable that concentrates on a few values requires less information to describe and therefore it is less uncertain with low entropy. On the other hand, if a random variable assumes a large number of values, it takes a long time to describe it, and therefore it is more uncertain with high entropy.

Let us denote by X the set of all possible values a random variable X can take. Denoting by p_X the probability density function of X, we can define the entropy of X as H(X):¹

$$H(X) = -\int p_X(x) \log(p_X(x)) \,\mathrm{d}x.$$
 (2.1)

If we have another random variable Y with probability density p_Y , and joint density $p_{X,Y}$, the joint entropy of X and Y denoted by H(X, Y) is given by

$$H(X,Y) = -\iint p_{X,Y}(x,y) \log(p_{X,Y}(x,y)) \,\mathrm{d}x \,\mathrm{d}y.$$
(2.2)

Building on this, the conditional entropy of Y given X is defined by

$$H(Y \mid X) = H(X, Y) - H(X)$$

= $-\iint p_{X,Y}(x, y) \log\left(\frac{p_{X,Y}(x, y)}{p_X(x)}\right) dx dy.$ (2.3)

¹ The base of the logarithm is not important. However, different bases result in different units. For example, \log_2 measures entropy in bits, \log_{10} measures entropy in dits, and $\log_e = \ln$ measures entropy in nats.

Finally, the mutual information between *X* and *Y* is calculated as follows:

$$I(X, Y) = H(X) - H(X | Y)$$

= $H(Y) - H(Y | X)$
= $H(X) + H(Y) - H(X, Y)$
= $\iint p_{X,Y}(x, y) \log \left(\frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)}\right) dx dy.$ (2.4)

Mutual information constructed as above is always nonnegative and equals zero if and only if X and Y are statistically independent in which case we have $p_{X,Y}(x, y) = p_X(x)p_Y(y)$. In order to obtain a distance statistic with certain desirable properties as described in Dionisio *et al* (2004), we will use a standardized measure defined by

$$d(X,Y) = 1 - \sqrt{1 - \exp(-2I(X,Y))}.$$
(2.5)

The distance measure in (2.5) is always in between 0 and 1. If d(X, Y) = 1, this implies that X contains no information on Y and vice versa. In this case the distance measure takes the maximum value, signifying that X and Y are far apart. If d(X, Y) = 0, there exists a perfect relationship between X and Y, or in other words, X and Y determine each other. Therefore, if there is a very close relationship between X and Y the distance measure is expected to be close to 0.

A number of algorithms exist for the estimation of mutual information. Meyer (2008) surveys the existing literature. We use the GNU R package "infotheo" by Meyer (2012) to carry out the computations. In particular, we chose options such that the mutual information is calculated from empirical probability distribution functions after discretization of continuous random variables related to asset returns using equal frequency bins.²

2.2 Minimum spanning trees

The mutual information distance (edge weights between nodes) between all assets (nodes of the network) constitute a fully connected network. For a universe of cardinality $n \in \mathbb{N}$, this network has n(n-1)/2. For example, when n = 50, we have 1225 pairs to think about. In order to reduce this complexity and reveal any potential clusters or potential patterns, we will give up some information by pruning a big portion of the network to come up with a tree. This will be a tree that connects every node to every other node and it connects every node in such a way that the total mutual information distance of the network is as small as possible with the condition that

www.risk.net/journal

² We experimented with a number of different ways of estimating the mutual information between assets to check the robustness of our findings. The outcomes did not differ significantly enough to report each and every experiment.

no cycles exist in the final tree. That is, we are after a compact representation of the initially complicated web of relationships with minimum possible information.

The mathematical representation of the MST is given as follows. Given a connected graph G = (V, E), with the set of nodes V, and the set of edges E, and weights d_e for all edges in E, we look for a spanning tree $G_T = (V_T, E_T)$ of minimum total distance. By "spanning" we mean $V_T = V$. A possible integer programming representation of the problem can be stated as follows. Let us represent by x_e the binary decision variable

$$x_e = \begin{cases} 1 & \text{if edge } e \in E_T, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \underset{x_e}{\text{minimize}} & \sum_{e \in E} d_e x_e \\ \text{subject to} & \sum_{e \in E} x_e = n - 1, \\ & \sum_{e \in (S,S)} x_e \leq |S| - 1, \ \forall S \subset V, \ S \neq \emptyset, \ S \neq V, \\ & x_e \in \{0,1\}, \ \forall e \in E, \end{aligned}$$

$$(2.6)$$

where (S, S) denotes all edges that go from a node in the set *S* to another node in the set *S*, and |S| means the cardinality of set *S*. The second constraint enforces that the edges in E_T cannot form cycles.

A desirable property of problem (2.6) is that if each edge weight is distinct then there is one and only one solution. This is a plausible assumption in real life since the probability of the exact equality between mutual information of pairs of assets is almost zero with distinct assets. Still, the cost of this compression is the loss of n(n-1)/2 - (n-1) = (n-1)(n-2)/2 edges, and the information accompanying them.

There are efficient heuristic algorithms that solve (2.6). Among them are Prim's and Kruskal's algorithms (see Cormen *et al* 2001). The most efficient designs of Prim's and Kruskal's algorithm have time complexities $O(n \log(n))$. The modern specialized solutions are known to improve the execution time to $O(\log(n))$ through parallel algorithms, as shown in Chong *et al* (2001).

2.3 Eccentricity

The eccentricity of a graph node $v \in V$ in a connected graph G is the maximum graph distance between v and any other node $u \in V$ of G. Intuitively, the eccentricity

Eccentricity in asset management 9



FIGURE 2 Eccentricity examples.

(a) A tree with two central nodes. (b) A tree with one central node.

measures how far a particular node is from the most distant border of the network. Therefore, nodes with low eccentricities are located toward the center of the network. The node (or nodes) with the minimum possible eccentricities are called central nodes. The eccentricity of a central node is called the radius of the network.

In Figure 2, nodes 1 and 2 both have eccentricities 2; therefore they are both central points and the radius of the graph is also 2. The remaining nodes have eccentricities equal to 3. In Figure 2, node 1 is the unique central point and its eccentricity is one. All other nodes have eccentricities equal to 2.

Figure 2 hints at an important condition we will be analyzing when we study the empirical properties of networks. Networks that are star-shaped, as in Figure 2, will be of certain interest. This is because in these types of graphs, nodes cluster around a single hub. When the nodes are assets, this so-called star topology corresponds to the maximum synchronization possible. Speaking generally to aid clarity, the information transfer between assets is bottle-necked through a single node, network is tight, and any shock can propagate through the rest of the network very quickly. These types of formations are potential harbingers of a catastrophe. On the contrary, a connected network that has a chain-like formation is the one with the least mutual information sharing. Information takes time to travel and a potential shock to one of the nodes does not necessarily result in a network wide shock. These formations correspond to calm or normal times.

The use of network centrality in determining systemic risk potential has lately been a hot research topic. For example, Lenzu and Tedeschi (2012) used network degree distribution, betweenness centrality and network diameter to characterize a smooth transition from a random topology to a star topology. Their results showed that during this transition only a few big financial institutions become central and carry capacity to transfer liquidity. This connectedness therefore leads to massive herding and consequently to more frequent insolvency-related bankruptcies. In a similar context, Hautsch *et al* (2014) use PageRank centrality to monitor companies' systemic importance by calculating the effect of the individual firm's contribution to systems' value-at-risk. Finally, the utilization of these network-related measures were the topic of the May 8–9, 2014 IMF conference "Interconnectedness: Building Bridges Between Policy and Research". As reported by Minoiu and Sharma (2014), during this conference it was concluded by Joseph Stiglitz of Columbia University that "the high degree of interconnectedness in the financial system facilitated the breakdown and became part of the problem".

In this paper, deviating from the liquidity-related bank borrowing/lending cases above, we will focus on return-based similarity distances to measure how close we are to a star-shaped formation. The reason behind our choice is twofold. First, while the above studies are important in analyzing the root causes, the underlying data may be lagged and hence may not be readily available to a practitioner such as a portfolio manager. Second, the systemic risk is probably more about the potential imminent contagion problem rather than who carries debt to whom. Even if two entities engage through healthy channels of lending and borrowing, they may be temporarily faced with defaults in the event of overreaction by market participants when everybody is running for the exit. In these cases, it is potentially better to infer the connectedness from return-based similarity measures. Therefore, we will attempt to understand the systemic risk potential by assessing how close the average network eccentricity is to 1 or whether the average eccentricity is increasing or decreasing. If it is close to 1 and is still decreasing, our hypothesis is that the network is transitioning to a star topology and a potential systemic shock is more probable.

3 DESCRIPTION OF THE DATA

In order to show the insensitivity of conclusions to the data used, we will work on two distinct sets of asset returns. The first set is a collection of asset classes including stocks, bonds and commodities. The second is a set of international equity indexes each representing a particular sector and industry pair.

Data is collected from Bloomberg and in each case we choose the periods in such a way that all assets have pricing data. In the case of global asset classes the period is from May 1990 to November 2013. In the case of sector/region pairs the period is from January 1996 to the end of November 2013.

While we used simple daily price returns for noncommodity assets, commodity returns are calculated from the most nearby futures contracts. A contract that has either a first notice date or last trade date in the next month is rolled forward. Due to possible time synchronization issues between the returns corresponding to different geographies, we take a two-day rolling average of returns, thereby introducing potential serial correlation to the returns we analyze, which we believe is not an essential drawback for the analysis we carry out. One-month local Libor rates are used to calculate excess returns for noncommodity indexes, commodity returns calculated from futures contract price changes are already excess returns. Table 1 on the next page and Table 2 on page 14 include detailed information and the legend of the asset identifiers used in the below study.

When calculating mutual information distances, we used a rolling one-year window with 252 business days. We repeated this exercise every month end, and in each iteration we used these distance matrixes to construct MSTs for that particular month. Next, we computed the eccentricity of each asset and recorded them for their time series analysis.

4 EMPIRICAL RESULTS

4.1 Compressing information by MSTs

Figure 3 on page 18 includes spanning trees as of November 2008 and November 2013. In part (a) of Figure 3 we can observe that during the financial crisis US stocks were located in the middle of the tree in between a group of emerging market (EM) countries and crude oil. One interpretation of this network formation is that US stocks, emerging market stocks and crude oil pretty much carried all the information that was needed during the crisis, perhaps related to the so-called global growth factor. The same is not true today as can be seen in part (b) of Figure 3. This current network looks more linear than the financial crisis network. Unlike before, commodities are not central. As a matter of fact, from this network commodities appear to carry less mutual information compared with stocks as commodities are located on the outer layers of the network, perhaps pricing their own idiosyncratic risks.

The sector/region networks also tell an interesting story. As we can imagine, financial stocks were very central during the crisis (see part (c) of Figure 3 on page 19). Today however, as can be seen from part (d), it is materials that are closer to the center of the network.

From parts (c) and (d) of Figure 3 on page 19 we can observe that there does not appear to be a sector-based clustering with some exceptions including materials

www.risk.net/journal

Class	Asset	Identifier	Bloomberg ticker
DM Stocks	MSCI USA Stock Index	US	MXUS Index
DM Stocks	MSCI Canada Stock Index	CA	MXCA Index
DM Stocks	MSCI Japan Stock Index	JP	MSDUJN Index
DM Stocks	MSCI Hong Kong Stock Index	HK	MSDUHK Index
DM Stocks	MSCI Australia Stock Index	AU	MSDUAS Index
DM Stocks	MSCI United Kingdom Stock Index	UK	MSDUUK Index
DM Stocks	MSCI Italy Stock Index	IT	MSDUIT Index
DM Stocks	MSCI Spain Stock Index	SP	MSDUSP Index
DM Stocks	MSCI Germany Stock Index	GE	MSDUGR Index
DM Stocks	MSCI France Stock Index	FR	MSDUFR Index
EM Stocks	MSCI Mexico Stock Index	MX	MXMX Index
EM Stocks	MSCI Brazil Stock Index	BR	MXBR Index
EM Stocks	MSCI Argentina Stock Index	AR	MXAR Index
EM Stocks	MSCI Chile Stock Index	CHL	MXCL Index
EM Stocks	MSCI Turkey Stock Index	TR	MSEUSTK Index
EM Stocks	MSCI Korea Stock Index	KO	MSEUSKO Index
EM Stocks	MSCI Philippines Stock Index	PH	MSEUSPHF Index
EM Stocks	MSCI Taiwan Stock Index	TA	MSEUSTW Index
EM Stocks	MSCI Thailand Stock Index	TH	MSEUSTHF Index
Bonds	Citigroup US GBI 7–10 year local currency	USB	SBUS70L Index
Bonds	Citigroup Canada GBI 7–10 year local currency	CAB	SBCD70L Index
Bonds	Citigroup UK GBI 7–10 year local currency	UKB	SBUK70L Index
Bonds	Citigroup Germany GBI 7–10 year local currency	EUB	SBDM70L Index
Bonds	Citigroup Japan GBI 7–10 year local currency	JPB	SBJY70L Index
Bonds	Citigroup Australia GBI 7–10 year local currency	AUB	SBAD70L Index
Energy	WTI crude oil	CL	CL1 Comdty
Energy	Brent crude oil	CO	CO1 Comdty
Energy	Natural gas	NG	NG1 Comdty
Energy	Heating oil	HO	HO1 Comdty
Energy	Gasoline	XB	XB1 Comdty
Energy	Gasoil	QS	QS1 Comdty

 TABLE 1
 Description of global assets. [Table continues on next page.]

Journal of Network Theory in Finance

www.risk.net/journal

Eccentricity in asset management 13

	Class	Asset	Identifier	Bloomberg ticker
Ν	Netals	Gold	GC	GC1 Comdty
Ν	Netals	Silver	SI	SI1 Comdty
Ν	/ letals	Platinum	PL	PL1 Comdty
Ν	Netals	Copper	HG	HG1 Comdty
Ν	Netals	Nickel	LN	LN1 Comdty
Ν	/ letals	Zinc	LX	LX1 Comdty
Ν	Netals	Aluminum	LA	LA1 Comdty
Ν	Netals	Lead	LL	LL1 Comdty
ŀ	Agriculture	Corn	С	C 1 Comdty
A	Agriculture	Soybean	S	S 1 Comdty
A	Agriculture	Wheat	W	W 1 Comdty
A	Agriculture	Kansas wheat	KW	KW1 Comdty
A	Agriculture	Soymeal	SM	SM1 Comdty
A	Agriculture	Bean oil	BO	BO1 Comdty
A	Agriculture	Cotton	СТ	CT1 Comdty
A	Agriculture	Sugar	SB	SB1 Comdty
ŀ	Agriculture	Cocoa	CC	CC1 Comdty
ŀ	Agriculture	Coffee	KC	KC1 Comdty
ŀ	Agriculture	Lean hogs	LH	LH1 Comdty
A	Agriculture	Feeder cattle	FC	FC1 Comdty
ŀ	Agriculture	Live cattle	LC	LC1 Comdty
	-			-

TABLE 1	Continued
---------	-----------

Commodity returns are calculated from rolling nearby contracts as they expire. The commodity indexes tabulated above cannot be used to calculate daily commodity returns.

and energy. Clustering seems to be more on the regional level. See, for instance, the current EM cluster in part (d). This suggests that it is perhaps more plausible for risk balanced global equity fund managers to do the risk budgeting on a regional level, rather than on a sector level.

4.2 Dynamics of the centrality of assets over time

In this subsection we will analyze the historical evolution of the average eccentricity of assets. In order to enhance the visual impact of the figures, we plot the deviation of an asset category's average eccentricity from the average eccentricity of all assets. Therefore, asset categories with negative deviation levels will imply that those asset categories are more central than positive deviation asset categories. The categorization

www.risk.net/journal

Sector	Region	ldentifier	Description	Bloomberg ticker
Finance	Europe	FNEU	MSCI EUR/FINANCE	MXEU0FN Index
Finance	North America	FNNA	MSCI N AMERICA/FINANCE	MXNA0FN Index
Finance	Pacific	FNPC	MSCI PACIFIC/FINANCE	MXPC0FN Index
Finance	Far East	FNFA	MSCI FAR EAST/FINANCE	MXFA0FN Index
Finance	Emerging markets	FNEM	MSCI EM/FINANCIALS	MXEF0FN Index
Energy	Europe	ENEU	MSCI EUR/ENERGY	MXEU0EN Index
Energy	North America	ENNA	MSCI N AMERICA/ENERGY	MXNA0EN Index
Energy	Pacific	ENPC	MSCI PACIFIC/ENERGY	MXPC0EN Index
Energy	Far East	ENFA	MSCI FAR EAST/ENERGY	MXFA0EN Index
Energy	Emerging markets	ENEM	MSCI EM ENERGY	MXEF0EN Index
Energy	Europe	MTEU	MSCI EUR/MATERIAL	MXEU0MT Index
Materials	North America	MTNA	MSCI N AMERICA/MATERIAL	MXNA0MT Index
Materials	Pacific	MTPC	MSCI PACIFIC/MATERIAL	MXPC0MT Index
Materials	Far East	MTFA	MSCI FAR EAST/MATERIAL	MXFA0MT Index
Materials	Emerging markets	MTEM	MSCI EM/MATERIAL	MXEF0MT Index

 TABLE 2
 Asset identifiers. [Table continues on next two pages.]

Journal of Network Theory in Finance

www.risk.net/journal

TABLE 2	Continued.				
	Sector	Region	ldentifier	Description	Bloomberg ticker
	Industrials	Europe	INEU	MSCI EUR/INDUSTRL	MXEU0IN Index
	Industrials	North America	INNA	MSCI N AMERICA/INDUSTRL	MXNA0IN Index
	Industrials	Pacific	INPC	MSCI PACIFIC/INDUSTRL	MXPC0IN Index
	Industrials	Far East	INFA	MSCI FAR EAST/INDUSTRL	MXFA0IN Index
	Industrials	Emerging markets	INEM	MSCI EM/INDUSTRL	MXEF0IN Index
	Consumer discretionary	Europe	CDEU	MSCI EUR/CONS DIS	MXEU0CD Index
	Consumer discretionary	North America	CDNA	MSCI N AMERICA/CONS DIS	MXNA0CD Index
	Consumer discretionary	Pacific	CDPC	MSCI PACIFIC/CONS DIS	MXPC0CD Index
	Consumer discretionary	Far East	CDFA	MSCI FAR EAST/CONS DIS	MXFA0CD Index
	Consumer discretionary	Emerging markets	CDEM	MSCI EM/CONS DISCR	MXEF0CD Index
	Consumer staples	Europe	CSEU	MSCI EUR/CON STPL	MXEU0CS Index
	Consumer staples	North America	CSNA	MSCI N AMERICA/CON STPL	MXNA0CS Index
	Consumer staples	Pacific	CSPC	MSCI PACIFIC/CON STPL	MXPC0CS Index
	Consumer staples	Far East	CSFA	MSCI FAR EAST/CON STPL	MXFA0CS Index
	Consumer staples	Emerging markets	CSEM	MSCI EM/CONSUMER STAPLES	MXEF0CS Index
	Health care	Europe	HCEU	MSCI EUR/HLTH CARE	MXEU0HC Index
	Health care	North America	HCNA	MSCI N AMERICA/HLTH CARE	MXNA0HC Index
	Health care	Pacific	НСРС	MSCI PACIFIC/HLTH CARE	MXPC0HC Index
	Health care	Far East	HCFA	MSCI FAR EAST/HLTH CARE	MXFA0HC Index
	Health care	Emerging markets	HCEM	MSCI EM/HLTH CARE	MXEF0HC Index

www.risk.net/journal

	Sector	Region	ldentifier	Description	Bloomberg ticker
	Information technologies	Europe	ITEU	MSCI EUR/INF TECH	MXEU0IT Index
	Information technologies	North America	ITNA	MSCI N AMERICA/INF TECH	MXNA0IT Index
	Information technologies	Pacific	ITPC	MSCI PACIFIC/INF TECH	MXPC0IT Index
	Information technologies	Far East	ITFA	MSCI FAR EAST/INF TECH	MXFA0IT Index
	Information technologies	Emerging markets	ITEM	MSCI EM/INF TECH	MXEF0IT Index
	Telecommunications	Europe	TCEU	MSCI EUR/TEL SVC	MXEU0TC Index
	Telecommunications	North America	TCNA	MSCI N AMERICA/TEL SVC	MXNA0TC Index
	Telecommunications	Pacific	TCPC	MSCI PACIFIC/TEL SVC	MXPC0TC Index
	Telecommunications	Far East	TCFA	MSCI FAR EAST/TEL SVC	MXFA0TC Index
	Telecommunications	Emerging markets	TCEM	MSCI EM/TEL SVC	MXEF0TC Index
	Utilities	Europe	UTEU	MSCI EUR/UTILITY	MXEU0UT Index
	Utilities	North America	UTNA	MSCI N AMERICA/UTILITY	MXNA0UT Index
	Utilities	Pacific	UTPC	MSCI PACIFIC/UTILITY	MXPC0UT Index
	Utilities	Far East	UTFA	MSCI FAR EAST/UTILITY	MXFA0UT Index
	Utilities	Emerging markets	UTEM	MSCI EM/UTILITY	MXEF0UT Index
All indexes e	cept European assets are in US dol	llars. Prior to the introduction	n of Euro, Deuts	che Mark is used to adjust returns to reflec	t exchange rate fluctuations.

TABLE 2 Continued.

Journal of Network Theory in Finance

www.risk.net/journal

definitions of global asset classes and sector/region equity indexes can be read from Table 1 on page 12 and Table 2 on page 14. In Figure 4 on page 20, we plotted these eccentricity deviations both for global assets and for sector/region equity indexes. In order to compare the centrality measures against the trends in the stock market, we included a visual overview of the performance of the S&P 500 price index.

In part (a) of Figure 4 on page 20 we can see that, prior to the financial crisis, developed market stocks and emerging market stocks were becoming more and more central. After the crisis, the centrality eased somewhat but, today, in contrast to emerging market stocks, developed market stocks look more central than before. Another interesting pattern on this figure is the recent central tendency of bonds. While still away from the center of the network, bonds are not as independent as they were during the period from late 2007 to late 2011. Perhaps this run toward the center has been a signal for the bond sell-off of 2013 all along. Finally, we also need to note that the energy and metals centralities tend to increase as they become less positive and eventually negative prior to market tops.

In part (b) of Figure 4 on page 21 we can observe that the financial sector was starting to become the center of the network prior to the 2008 crisis, and has remained very central after the crisis. Similar notable deviations can be observed in information technology and telecommunications stocks prior to the dotcom bubble. As of today, materials tend to be in the center of the network. The utilities and health care sectors have consistently found themselves on the outer layers of the network; perhaps proving the widespread claim that they are the bond-like investments in the equity domain pricing independent factors.

4.3 Dynamics of the topology of the network for market timing

In this section the goal is to see whether certain trends in average network eccentricity precede market downturns. Our hypothesis is that networks that are star-shaped are more likely to precede systemic events than chain-like graphs. This is because star-shaped networks have maximum information interconnectedness. Any potential shock to a particular node can simultaneously get translated into a system-wide catastrophic event.

In order to measure the star-likeness of the network we will look at two indicators. The first is the level of the average network eccentricity, which is simply the crossaverage eccentricity of the set of all nodes at a particular time. The second indicator is the change in this level of average eccentricity indicator. To measure the change, we will subtract the twelve-month moving average of level from its six-month moving average. A positive change indicates that the network is becoming more chain-like, whereas a negative change is an indication of star-likeness.

www.risk.net/journal



FIGURE 3 Current and recession period MSTs. [Figure continues on next page.]

(a) MST of global asset classes on October 31, 2008. (b) MST of global asset classes on October 31, 2013.

Journal of Network Theory in Finance

www.risk.net/journal

Eccentricity in asset management 19



FIGURE 3 Continued.

(c) MST of MSCI sector/region indexes on October 31, 2008. (d) MST of MSCI sector/region indexes on October 31, 2013. Data from Bloomberg and calculations from Neuberger Berman Quantitative Investment Group.

www.risk.net/journal



FIGURE 4 Deviation of asset eccentricity from average network eccentricity of MSCI region/sector pairs over time. [Figure continues on next page.]

(a) Centrality of global asset classes over time.

4.3.1 History of the network topology

Figure 5 on page 23 includes levels and changes in average network eccentricity both in the case of global asset classes and in the case of sector/region indexes. To indicate market tops and bottoms, S&P 500 price series is scaled to fit each figure part.

Of interest in the past two decades are the dotcom bubble and the global financial crisis. Looking at parts (a) and (c) of Figure 5 on page 23 we see that, while both





(b) Centrality of MSCI region/sector pairs over time. S&P 500 price series, not to be confused with S&P 500 centrality series, is scaled to fit the figure. Data from Bloomberg and calculations from Neuberger Berman Quantitative Investment Group.

exhibiting tightness with low eccentricity readings, the equity sector/region based average network eccentricity level gave a clearer warning signal prior to the dotcom bubble bursting. Perhaps this is because the dotcom bubble was more of a regional bubble than a large scale global event. Second, prior to the 2008 sell-off we see that both indicators started to decline. In this case the decline in the global asset classes

www.risk.net/journal

network eccentricity level is starker than the equity counterpart. The decline goes until 2010 in the case of sector/region based eccentricity level and until early 2011 in the case of global assets. These observations show that the average network eccentricity level does a better job in timing the peaks than troughs.

Parts (b) and (d) of Figure 5 on the facing page tell a similar story. The market highs in both sell-offs were preceded by a negative regime in average eccentricity trends. These figures also show a drawback of the market timing function of eccentricity measures. There are a number of false indications. For example, the false negative in late 1997 due to the Asian crisis did not result in an imminent market sell-off in the US. Again the false negative in the late 2003/early 2004 global asset eccentricity change indicator did not lead to a bear market. This negative signal that appears to be mostly driven by stocks and bonds (as can be seen in part (a) Figure 4 on page 20) also coincided with the period when Federal Reserve ended its expansionary program by increasing rates in June 2004. Perhaps it is the Federal Reserve action in this period that prevented a potential sell-off.

4.3.2 Network topology and subsequent market returns

To tie up the loose ends in the previous section we calculated the average network eccentricity behavior prior to market crashes. Without loss of generality, we define a crash month to be a month in which S&P 500 returned less than its historical 5th percentile. For each of these months we recorded the level of average network eccentricity prior to, during and after the crash. We record from six months before to six months after. Next, we take the average of network eccentricities in each of these before and after months across all the crash months.

Parts (a) and (c) of Figure 6 on page 24 show the average level of network eccentricity around crashes first in global asset classes and the second in sector/region indexes. Both figures indicate that the average level of eccentricity tends to decline prior to a crash. Keeping an eye on the level may therefore be useful. After the crash the level rebounds in global asset classes but there is no significant change in sector/region pair.

The systemic event timing value can perhaps be better observed in conditional return densities. To this end, we classified next month's returns into two categories. Category "decreasing" includes those months which were preceded by a negative trend in the level of eccentricity as measured by the difference between six-month and twelve-month moving averages of level. Similarly, the category "increasing" includes those months that were preceded by a positive trend in the level of eccentricity. In other words, the "increasing" class represents those periods that were preceded by a network that is more chain-like and the "decreasing" periods were preceded by networks that were more star-like.



FIGURE 5 History of the level and the change in the level of network eccentricities in global asset classes and in MSCI region/sector indexes.

(a) Level of the network eccentricity in global asset classes. (b) Change in the level of the network eccentricity in global asset classes. (c) Level of the network eccentricity in MSCI region/sector pairs. (d) Change in the level of the network eccentricity in MSCI region/sector pairs. S&P 500 price series is scaled to fit. In each case, change in level is calculated by subtracting the twelve-month moving average of levels from six-month moving average of levels. Data from Bloomberg and calculations from Neuberger Berman Quantitative Investment Group.

Parts (b) and (d) of Figure 6 on the next page both show that there is a clear distinction between the densities of returns when they are conditioned on the network eccentricity level change. In both cases the "decreasing" density is more left skewed and significantly more leptokurtic (with fatter left tails). Of course we should not take this as a home run because positive returns are also likely in "decreasing" eccentricity regimes. It is probably safer to claim that while a "decreasing" network eccentricity is a necessary condition for a sell-off it is definitely not a sufficient condition.

www.risk.net/journal





(a) Average network eccentricity around crashes in global asset classes. (b) Density plot of S&P 500 price returns conditional on the state of network eccentricity trends in global asset classes. (c) Average network eccentricity around crashes in MSCI region/sector pairs. (d) Density plot of S&P 500 price returns conditional on the state of network eccentricity trends in MSCI region/sector pairs. A period in which the S&P 500 loses more than its historical 5th percentile of returns is called a crash period. Decreasing is the regime in which in the prior month six-month moving average of the network eccentricity is below the twelve-month moving average. Increasing is the regime in which in the prior month six-month moving average of the network eccentricity plot indicates that decreasing trends in network eccentricity may indicate a heightened probability of tail events. Data from Bloomberg and calculations from Neuberger Berman Quantitative Investment Group.

4.4 Centrality versus return

While in this introductory paper our aim is not to fish for a yet another asset pricing factor and an accompanying theory to support it, it is natural to make a link between CAPM and asset centrality. In CAPM assets that are more exposed to the market, as measured by their beta, in equilibrium, need to earn higher expected risk premiums. The key concept here is exposure to the market, which is the center of gravity of the capitalization weighted investment universe. In a similar manner, can assets that are more exposed to the center of an asset network earn higher expected premiums?





(a) Average asset returns versus average centrality in global asset classes. (b) Average asset returns versus average centrality in MSCI region/sector pairs. In both cases, the average return of an asset increases as the asset becomes more central. Data from Bloomberg and calculations from Neuberger Berman Quantitative Investment Group.

www.risk.net/journal

To answer this question, we compute the historical eccentricity of each asset by averaging the historical eccentricities in each period and using them as independent variables in regressions against historical average returns. To check robustness, we carried out these regressions in whole sample, and in two half subsample periods. We also repeated this exercise in both global asset classes and in sector/region equity indexes.

Parts (a) and (b) of Figure 7 on the preceding page empirically support the claim that central assets tend to earn higher returns. In both cases, in each subsample, and in each whole sample, the regression lines are negatively sloping. It is also clear from the fitted error bounds that the variation tends to be explained better in average eccentricity levels than in extreme centrality cases. This is due to the lack of bulk data in the extremal parts of the network. In other words, while there are many noncentral and nonboundary assets, there are fewer central and boundary assets.

This analysis is by no means a conclusion that asset centrality is an economic risk factor. This is despite the fact that recent research has started to tend in this direction. See, for example, Ozsoylev and Walden (2011) and Buraschi and Porchia (2012) for similar empirical findings. Further research is necessary to link the centrality to a stochastic discount factor. Without such a link, we can at most call centrality a statistical factor similar to principle components of a covariance matrix.

4.5 Asset allocation with eccentricity

In this final section, we would like to carry out backtests to assess the historical economic value added by eccentricity as a conditioning variable. To do that we devised three benchmarks. The first one is the equal weighted portfolio of all assets. The second is the volatility parity, which weights assets according to the inverse of their historical rolling one-year volatilities. The third one is what we call the eccentricity budgeting, which budgets risks proportional to the eccentricity of assets. In other words, if volatility parity weights an asset $i \in \{1, 2, ..., n\}$ in a universe with $n \in \mathbb{N}$ assets proportional to the inverse of the volatility $w_i \sim 1/\sigma_i$, then the eccentricity budgeting weights the same asset inversely proportional to its eccentricity ε_i by $w_i \sim 1/\varepsilon_i \times 1/\sigma_i$. All portfolios are scaled to have sum of weights equal to 1 to restrict leverage.

Figure 8 on the facing page includes the backtest performance of each portfolio construction method. Parts (a) and (b) include the cumulative returns, monthly returns, and drawdowns and Table 3 on page 28 includes average performance statistics. In each domain, we see that volatility parity improves on equal weighting, and eccentricity budgeting improves on volatility parity. The improvements are based on return to risk ratios including Sharpe ratio, mean over value-at-risk, mean over conditional value-at-risk, and Calmar ratio (mean/maximum drawdown).



FIGURE 8 Historical performance of eccentricity budgeting compared to volatility parity and equal weighted portfolios.

(a) Backtest of eccentricity budgeting in global asset classes. (b) Backtest of eccentricity budgeting in MSCI region/sector pairs. Data from Bloomberg and calculations from Neuberger Berman Quantitative Investment Group.

Journal of Network Theory in Finance

www.risk.net/journal

TABLE 3Data for Figure 8.

	Eccentricity budgeting	Volatility parity	Equal weighted
Return (cumulative, excess)	116.63	83.97	63.61
Return (annualized, excess)	3.50	2.75	2.21
Volatility (annualized)	8.04	7.65	12.27
Sharpe ratio (annualized)	0.43	0.36	0.18
VaR (95%) (monthly)	3.89	3.95	6.43
CVaR (95%) (monthly)	6.60	9.17	14.20
Maximum drawdown	34.21	33.06	49.31
Annual return/monthly CVaR (95%)	0.53	0.30	0.16
Calmar ratio	0.10	0.08	0.04

(b) Performance of eccentricity budgeting in MSCI region/sector pairs

	Eccentricity budgeting	Volatility parity	Equal weighted	
Return (cumulative, excess)	68.68	52.75	40.11	
Return (annualized, excess)	2.98	2.40	1.91	
Volatility (annualized)	13.77	14.90	15.94	
Sharpe ratio (annualized)	0.22	0.16	0.12	
VaR (95%) (monthly)	7.05	7.76	8.29	
CVaR (95%) (monthly)	10.43	12.00	12.90	
Maximum drawdown	46.84	51.20	54.29	
Annual return/monthly CVaR (95%)	0.29	0.20	0.15	
Calmar ratio	0.06	0.05	0.04	

Although we did not include transaction costs in these backtests, we still believe eccentricity budgeting may be beneficial as the returns to risk ratios are different enough from their naive counterparts. Especially allowing for leverage, eccentricity budgeting may perhaps be an input for risk-balanced investment frameworks.

5 CONCLUSION

In this paper we studied the elusive concept of synchronization in asset returns. To avoid any ambiguity about its definition, we started with an assumption that synchronization meant increasing mutual information between assets. We measured the closeness of assets according to their mutual information distance. This distance led
to a fully connected graph of mutual information relationships. To reduce the clutter in these graphs, we utilized MSTs. On these trees, we defined a centrality measure called eccentricity which was defined as the shortest distance of a node from the farthest out node in the graph.

We carried out empirical studies on two distinct domains. First, in a set of global assets including stocks, bonds and commodities and second in a set of sector/region equity indexes. The combined time period of these studies was from the early 1990s to late 2013.

We analyzed the centrality of assets over time and observed that developed market stocks and emerging market stocks were mostly central with low eccentricities. However, during the crisis energy and metal commodities became central as well. In the sector/region domain, we saw how the financial sector trended toward centrality prior to the 2008 meltdown. Similar patterns were observed in IT and telecom sectors prior to the dotcom bubble burst.

We looked at the market timing ability of the average network eccentricity level. We saw that the network eccentricity tended to decline prior to market crashes and densities of returns can be statistically distinguished based on prior months' average network eccentricity trend. This is especially so in the left tails. However, we also noted that the eccentricity based market timing indicators can give false alarms.

The asset pricing implications of eccentricity were also studied. Historical regressions showed that assets with central eccentricity earned higher returns on average. Analysis in different subperiods supported the robustness of this empirical finding. However, due to the lack of theoretical background, we deduced that at this stage eccentricity can only be classed as a statistical risk factor.

Finally, we carried out backtests to assess the potential economic value added by eccentricity information. We saw that in each of the domains we considered, eccentricity-based risk budgeting portfolios improved the return to risk ratios compared to volatility parity and equally weighted portfolios.

It may be advisable to keep an eye on the centrality of certain assets and the average centrality of the network by comparing them to their historical episodes. We can then look at what subsequent events followed these similar periods. In the portfolio context, we can monitor the portfolio weighted average of the eccentricities of all assets. If the portfolio eccentricity is too low, perhaps it is advisable to shift some of the central weights to the outer layers of the network for diversification purposes.

In sum, while still in its infancy, information-theoretic networks can be a useful instrument in a portfolio manager's toolbox. As more data becomes available, and is cheaper in higher frequencies, the applicability and precision of the predictions of this framework may well increase. Perhaps until then alpha may continue to grow on trees.

DECLARATION OF INTEREST

This material is provided for informational purposes only and nothing herein constitutes investment, legal, accounting or tax advice, or a recommendation to buy, sell or hold a security. Information is obtained from sources deemed reliable, but there is no representation or warranty as to its accuracy, completeness or reliability. All information is current as of the date of this material and is subject to change without notice. Any views or opinions expressed may not reflect those of the firm as a whole. Neuberger Berman products and services may not be available in all jurisdictions or to all client types. Investing entails risks, including possible loss of principal. Investments in alternative products, including hedge funds and private equity are speculative and involve a higher degree of risk than more traditional investments. Indexes are unmanaged and are not available for direct investment. Past performance is no guarantee of future results.

REFERENCES

- Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2010). Measuring systemic risk. Working Paper 1002, Federal Reserve Bank of Cleveland. URL http:// EconPapers.repec.org/RePEc:fip:fedcwp:1002.
- Adrian, T., and Brunnermeier, M. K. (2011). CoVaR. Working Paper 17454, National Bureau of Economic Research, October. URL: www.nber.org/papers/w17454.
- Barigozzi, M., and Brownlees, C. (2013). Nets: network estimation for time series. Working Paper 723, Barcelona Graduate School of Economics, October. URL: http://ideas.repec.org/p/bge/wpaper/723.html.
- Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* **104**(3), 535–559.
- Bisias, D., Flood, M., Lo, A., and Valavanis, S. (2012). A survey of systemic risk analytics. US Department of Treasury, Office of Financial Research (0001).
- Boginski, V., Butenko, S., and Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis* 48(2), 431–443.
- Buraschi, A., and Porchia, P. (2012). Dynamic networks and asset pricing. In *AFA* 2013 San Diego Meetings Paper.
- Chong, K. W., Han, Y., and Lam, T. W. (2001). Concurrent threads and optimal parallel minimum spanning trees algorithm. *Journal of the ACM* **48**(2), 297–323.
- Cormen, T. H., Stein, C., Rivest, R. L., and Leiserson, C. E. (2001). *Introduction to Algorithms*, 2nd edition. McGraw-Hill Higher Education.
- Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience.
- Diebold, F. X., and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal* **119**(534), 158–171.

- Dionisio, A., Menezes, R., and Mendes, D. A. (2004). Mutual information: a measure of dependency for nonlinear time series. *Physica A: Statistical Mechanics and Its Applications* **344**(1), 326–329.
- Engle, R., and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics* **30**(2), 212–228.
- Fuld, R. S. (2008). Lehman's Fuld says "worst is behind us" in crisis. *Bloomberg*, April 15. URL: www.bloomberg.com/apps/news?pid=newsarchive&sid=a5JCJBH76_VU.
- Hage, P., and Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks* **17**(1), 57–63.
- Hautsch, N., Schaumburg, J., and Schienle, M. (2014). Financial network systemic risk contributions. *Review of Finance*, DOI:10.1093/rof/rfu010. URL: http://rof.oxfordjournals.org/content/early/2014/03/24/rof.rfu010.abstract.
- Hiemstra, C., and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price–volume relation. *Journal of Finance* **49**(5), 1639–1664.
- Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., and Ben-Jacob, E. (2010). Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PloS One* 5(12), e15032.
- Kenett, D. Y., Preis, T., Gur-Gershgoren, G., and Ben-Jacob, E. (2012). Dependency network and node influence: application to the study of financial markets. *International Journal of Bifurcation and Chaos* 22(7), 1250181/1–14.
- Klößner, S., and Wagner, S. (2014). Exploring all VAR orderings for calculating spillovers? Yes, we can! A note on Diebold and Yilmaz (2009). *Journal of Applied Econometrics* **29**(1), 172–179.
- Kolanovic, M., Silvestrini, D., Lee, T. S. K., and Naito, M. (2011). Rise of cross-asset correlations. JP Morgan Global Equity Derivatives & Delta One Strategy, May. URL: www.cboe.com/Institutional/JPMCrossAssetCorrelations.pdf.
- Koschützki, D., and Schreiber, F. (2004). Comparison of centralities for biological networks. In *Proceedings of the German Conference on Bioinformatics (GCB'04)*, Volume P-53 of LNI, p. 53.
- Koschützki, D., and Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regulation and Systems Biology* **2**, 193–201.
- Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical Review E* **69**, 066138.
- Kritzman, M., Li, Y., Page, S., and Rigobon, R. (2010). Principal components as a measure of systemic risk. SSRN eLibrary.
- Lautier, D., and Raynaud, F. (2013). Systemic risk and complex systems: a graphtheory analysis. In *Econophysics of Systemic Risk and Network Dynamics*, Abergel, F., Chakrabarti, B. K., Chakraborti, A., and Ghosh, A. (eds), pp. 19–37. New Economic Windows, Springer.
- Lenzu, S., and Tedeschi, G. (2012). Systemic risk on different interbank network topologies. *Physica A: Statistical Mechanics and Its Applications* **391**(18), 4331–4341.
- Majdandzic, A., Podobnik, B., Buldyrev, S. V., Kenett, D. Y., Havlin, S., and Stanley, H. E. (2014). Spontaneous recovery in dynamical networks. *Nature Physics* **10**(1), 34–38.
- Mantegna, R. N. (1999). Hierarchical structure in financial markets. European Physical Journal B: Condensed Matter and Complex Systems 11(1), 193–197.

www.risk.net/journal

- 32 H. Kaya
 - Meyer, P. E. (2008). Information-theoretic variable selection and network inference from microarray data. PhD thesis, Université Libre de Bruxelles, Brussels, Belgium.
 - Meyer, P. E. (2012). infotheo: information-theoretic measures. R package, version 1.1.1. URL: http://CRAN.R-project.org/package=infotheo.
 - Minoiu, C., and Sharma, S. (2014). Financial networks key to understanding systemic risk. *IMFSurvey Magazine: IMF Research*, May 28. URL: www.imf.org/external/pubs/ ft/survey/so/2014/res052314a.htm.
 - Moon, Y.-I., Rajagopalan, B., and Lall, U. (1995). Estimation of mutual information using kernel density estimators. *Physical Review* E 52, 2318–2321.
 - Ozsoylev, H. N., and Walden, J. (2011). Asset pricing in large information networks. *Journal of Economic Theory* 146(6), 2252–2280.
 - Rothman, P. (1999). Nonlinear Time Series Analysis of Economic and Financial Data, Vol. 1. Springer.
 - Sandoval, L. (2014). Structure of a global network of financial companies based on transfer entropy. *Entropy* **16**(8), 4443–4482.
 - Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization* **75**(1), 40–58.
 - Walters-Williams, J., and Li, Y. (2009). Estimation of mutual information: a survey. In *Rough Sets and Knowledge Technology*, Wen, P., Li, Y., Polkowski, L., Yao, Y., Tsumoto, S., and Wang, G. (eds), pp. 389–396. Lecture Notes in Computer Science, Vol. 5589. Springer.

www.risk.net/journal



Research Paper

Emergence of the EU corporate lending network

Grzegorz Hałaj, Urszula Kochańska and Christoffer Kok

European Central Bank, Sonnemannstrasse 22, D-60314 Frankfurt am Main, Germany; emails: grzegorz.halaj@ecb.int, urszula.kochanska@ecb.int, christoffer.kok@ecb.int

(Received October 1, 2014; revised December 19, 2014; accepted February 2, 2015)

ABSTRACT

This paper uses network formation techniques based on a theoretical framework developed by Hałaj and Kok to construct networks of lending relationships between a large sample of banks and nonbanks in the EU. The model provides an assessment not only of how banks are directly related to each other in the interbank market, but also how they may be indirectly related (due to common exposures) via their corporate lending relationships. We illustrate how the model can be used to conduct counterfactual simulations of the contagion effects arising when individual – or groups of – banks and firms are hit by shocks. This could allow policy makers to gauge specific vulnerabilities in the financial system evolving around the lending relationships between banks and their (corporate) borrowers. Furthermore, we show that the modeling framework can be used by micro- and macroprudential authorities to analyze the impact of varying banks' large exposure limits as a way to mitigate contagion effects within and beyond the financial system.

Keywords: interbank network; contagion; corporate lending; portfolio choice; exposure limits.

Corresponding author: Grzegorz Hałaj

Print ISSN 2055-7795 | Online ISSN 2055-7809 Copyright © 2015 Incisive Risk Information (IP) Limited

1 INTRODUCTION

This paper uses network formation techniques based on the theoretical framework of Hałaj and Kok (2014) to construct networks of lending relationships between a large sample of banks and nonbanking corporations in the EU.¹ Networks of bankfirm lending relationships provide an alternative approach to studying real-financial linkages, which takes into account the heterogeneous characteristics of individual banks and firms on the propagation of shocks between the financial sector and the real economy. One particular strength of the model is related to the fact that the proposed framework provides an assessment not only of how banks are directly related to each other in the interbank market but also how they may be indirectly related (due to common exposures) via their corporate lending relationships. The model can be used to conduct counterfactual simulations of the contagion effects arising when individual - or groups of - banks and firms are hit by shocks. This could allow policy makers to gauge specific vulnerabilities in the financial system evolving around the lending relationships between banks and their (corporate) borrowers. Furthermore, we show that the modeling framework can be used by micro- and macroprudential authorities to analyze the impact of varying banks' large exposure limits as a way to mitigate contagious effects within and beyond the financial system.

The monitoring and assessment of vulnerabilities related to bank-firm lending relationships are important for several reasons.

It is well-known that banks play a key role in intermediating savings into productive uses, especially via the financing of the nonfinancial corporate sector. The lending relationships between banks and firms that are created in this intermediation process thus contribute to the smooth functioning of the economy be reallocating savings (typically from households) to entities with a financing need (such as, nonfinancial corporations). This is especially the case in bank-based financial systems, such as the euro area and Japan, where nonfinancial corporations are largely dependent on bank financing whereas more market-based sources of financing are less prevalent. However, even in more market-based financial systems, such as the US and UK, bank lending plays a nonnegligible role in the economy; not least for the financing of small and medium-sized enterprises.

The bank–firm lending relationships furthermore play an important role in the monetary policy transmission mechanism; especially via the so-called bank lending channel where the transmission of monetary policy impulses depends on the conditions of the banking sector (see, for example, European Central Bank 2008). In the euro area, the importance of the bank lending channel has been particularly obvious during the financial crisis and the euro area sovereign debt crisis where the channels

¹ The sample also includes Norway, but all other banks and firms are from the EU.

of transmissions became clogged due to serious impairment of banks' balance sheets as well as (in some instances) country-specific problems related to the weak fiscal position of the sovereign (see, for example, European Central Bank 2012).

It is also important from a prudential perspective to monitor bank-firm lending relationships.

First, microprudential supervisors have a keen interest in surveying the bank's direct credit exposures to the nonfinancial corporate sector, as shocks hitting the latter are likely to have negative repercussions on the balance sheet of the bank. Such considerations may lead to prudential requirements to the bank's loan loss provisioning amounts and to reassessments of the appropriateness of their credit risk management models.

Second, however, banks (and their supervisors) may often overlook the macroprudential implications arising due to the fact that balance sheet impairments to individual banks can also occur due to indirect contagion effects resulting from interbank linkages that could potentially amplify shocks to individual banks' corporate credit portfolios. For example, shocks to the creditworthiness of nonfinancial corporations that may impair a firm (or a group of firms) to repay the debt owed to a bank could have broader implications as other banks may also have exposures to those firms ("common exposures"). Moreover, if the resulting impairments to its credit portfolio are large enough, it could also directly hamper the bank's ability to meet its obligations in the interbank market or at least induce it to withdraw funding to other banks, which could therefore trigger contagion effects within the interbank market that ultimately may lead other banks to reduce lending to their corporate borrowers.

Taking a macroprudential perspective, this paper presents a framework that allows for capturing the contagion effects that may arise not only due to the interbank linkages but also taking into account the many linkages that exist between banks and the nonbanking corporate sector. Importantly, for the formation of the networks we assume that banks (and firms) dynamically optimize their lending and funding decisions subject to regulatory and economic constraints and nonbank firms try to optimize structures of their financing received in the form of bank loans.

For this purpose, we present an agent-based network model of the banking sector extended with a network of bank–firm relationships. The modeling framework is based on Hałaj and Kok (2014). It is based on aggregate data on banks' corporate loan portfolios, interest rates of corporate loans and some proxy for potential lending relationship for different segments of the market. The model is based on a constrained optimization framework whereby firms' demand for loans from individual banks is driven by optimizing decisions related to their external financing needs, a constrained number of desired bank lending relationships (based on micro-level credit register data), geographical proximity to banks in the system and the perceived default risk of those banks. Banks are similarly assumed to be allocating their provision of loans to

the corporate sector and on the interbank market subject to constrained optimization. On the one hand, banks' decisions to provide loans to corporates are determined by risk-return considerations related to individual corporate borrowers and subject to regulatory (capital) constraints. The interbank market activity (lending and borrowing) is assumed to be driven mainly by liquidity management considerations, partly in response to fluctuations in excess liquidity as corporate lending evolves over time. The banks' interbank funding decisions are assumed to be a function of perceived counterparty credit risk to minimize potential rollover risk as well as geographical proximity. In contrast to the corporate sector's bank financing decision, the number of potential interbank funding counterparts is assumed to be substantially larger.

The paper aims to improve our understanding of the linkages between banks and the real economy while accounting for the heterogeneous behavior of individual agents and how the dynamic interactions with other agents may affect the overall propagation of shocks to the economy. The network formation approach presented in the paper thus allows us to assess the risks stemming from interconnectedness between banks and between banks and the corporate sector. The modeling framework can also be used to assess how different macroprudential policy choices (such as as changing bank capital requirements) can affect the contagion risks inherent in the bank–firm network.

The applicability of the model for policy analysis is illustrated using a number of counterfactual simulations. The main findings of the paper are the following.

- A similar-sized shock to creditworthiness across all nonbank financial sectors can imply markedly different effects in terms of the contagion losses that individual sectors inflict upon the banking sector, and also differences across countries are notable. The size of contagion losses is likely to depend on economic and financial structures and on the concentration and riskiness of specific sectors in banks' portfolios.
- Shocks to the creditworthiness of one domestic nonbank corporate sector, in addition to the impact on the domestic banking sector, may create material cross-border contagion effects on banking sectors in other European countries. Specific sectors' ability to create substantial cross-border contagion effects hinges on their international activities and on the cross-border links of the domestic banking sector.
- Contagion effects from nonbank corporations to the banking sector can also feed back to the real economy via banks' reactions to the initial shock which may induce them to cut back on funds provided to the corporate sector. This

points to the importance of macro feedback loops when considering the close interlinkages between banks and the real economy.

• Shocks to the banking sector (eg, as typically analyzed in stress tests) can have important second-round effects on the nonbank corporate sector via the bank–firm lending relationships generated by the model. We also show that such contagion risk can be at least partly stemmed by using micro- and macroprudential policy tools, such as large exposure limits and changes in risk weights.

The paper contributes to the literature by extending the standard interbank network to also encompass the real economy by adding a network layer consisting of nonfinancial corporations. To our knowledge, this is one of the first examples of a firm-level financial network model incorporating links to the real economy in a "network fashion". A few other recent studies also explore bank–firm relationships using network techniques, such as de Castro Miranda and Tabak (2013) using Brazilian data, Masi and Gallegati (2012) using Italian data and Aoyama *et al* (2009) and Gallegati *et al* (2010) using Japanese data.

Furthermore, our model builds on agent-based approaches whereby the financial-toreal networks are formed via banks' (and firms') optimizing behavior. Our approach thus goes beyond the standard financial network literature, which has traditionally focused in particular on the interbank market/payment system as a source of contagion but which has not focused on explaining how the interbank networks emerge or how they may be affected by changing financial conditions (see, for example, Allen and Babus 2009; Allen and Gale 2000; Iori et al 2008; Nier et al 2007). For empirical interbank network studies using overnight interbank transactions data at national level, see, for example, Furfine (2003), Upper and Worms (2004), Boss et al (2004), van Lelyveld and Liedorp (2006) and Soramaki et al (2007). Instead, our paper is more closely related to a more recent strand of the financial network literature that uses portfolio optimization (see, for example, Georg 2013), game theory (see, for example, Acemoglu et al 2013; Aldasaro et al 2014; Blasques et al 2014; Bluhm et al 2013; Cohen-Cole et al 2011), stochastic games and matching problems (see, for example, Chen and Song 2013; Duffie and Sun 2012; Eisenschmidt and Tapking 2009; Jackson and Watts 2010) and agent-based approaches to address overly complex equilibria (see, for example, Grasselli 2013; Hałaj 2012; Hałaj and Kok 2014; Markose 2012; Montagna and Kok 2013). Castiglionesi and Lavarro (2011) develop a network formation model with micro-founded bank behavior that is related to our study (see also Babus 2011; Castiglionesi and Wagner 2013). Our modeling approach is also related to studies such as t'Veld et al (2014), who model interbank network formation via core-periphery structures as a function of the banks' intermediation activity (see also Manea 2014).

The model presented in this paper is set up in a way that it can easily be linked to traditional bank stress testing analysis to study the effects of adverse scenario shocks affecting the banks or the nonbank firms, respectively. It is thus also related to interbank network models used to augment stress testing frameworks, such as, for example, Canedo and Martínez-Jaramillo (2009), Lopez-Castañon *et al* (2012) and Hałaj and Kok (2013).

Finally, our paper is also related to the extensive empirical literature on bank–firm lending relationships, which points to the importance of relationships in overcoming asymmetric information problems but also highlights potential borrower capture and the effects it may have on access to bank finance.² Of particular relevance to our paper is the finding in this strand of the literature that most firms tend to operate with only a small number of bank lending relationships. This finding is corroborated by data from euro area credit registers data that we employ to help generate the bank–firm lending relationships in our model.

The paper is structured as follows. Section 2 presents the data used to generate the interbank and bank–firm networks. Section 3 describes the network formation model, while in Section 4 we exploit the model to run various simulations that illustrate how different contagion effects can be captured. It is also illustrated how prudential policy tools can be employed to potentially mitigate the contagion effects inherent in the interconnected bank–firm networks. Section 5 concludes.

2 DATA

Our bank–firm network model is meant to mirror the bank-to-bank and bank-to-firm interlinkages between corporations in the EU countries. The country coverage is determined by the bank sample which corresponds to the EU countries with banking groups participating in the EBA 2011 stress test exercise. These banks have been subject to regular EBA disclosures that provide a level of granular bank data suitable for generating interbank networks (as shown in Hałaj and Kok (2013)). The input data for the model consists of largely public data on banks and firms other than banks, while also taking recourse to some confidential credit register-based information about the number of firms' bank relationships. Table 1 on page 40 provides an overview of the data dimensions, while Table 2 on page 42 and Table 3 on page 46 provide key descriptive statistics.

2.1 Firms other than banks

Data on individual firms is based on official statistics and market data. The sample of nonbank firms is derived from the members of the benchmark equity indexes in the EU

² For a few seminal studies see, for example, Petersen and Rajan (1994), Berger and Udell (1995), Boot (2000), Boot and Thakor (2000) and Angelini *et al* (1998).

countries listed in the footnote to Table 1 on the next page. In total, market price and balance sheet data for around 900 firms from the blue chip indexes of the major stock exchanges in the EU have been collected from Bloomberg. Figure 1 on page 49 shows the aggregate total assets at year-end 2012 for the firms in the sample broken down by country and by industry sector classification. The in-sample companies constitute more than 75% of the total listed nonbank firms in the analyzed countries in terms of total assets. Table 2 on page 42 and Table 3 on page 46 present the dispersion of some key parameters of the firms in the sample. It can be observed that the sample is quite heterogenous in term of the firms' total assets (median size company has $\in 1.7$ billion total assets whereas the standard deviation of the total assets of firms amounts to \in 70.6 billion) and the volume of borrowing from banks. In addition to the total asset figures, country and NACE sector codes, information on the companies' total liabilities, total equity and measures of credit risks were collected. As regards credit risk measures, the CDS spreads on senior debt with five years' maturity and long-term issuer ratings by Moody's, Fitch and S&P were collected. Where CDS information was not available, the average expected default frequencies (from Moody's KMV) within one year for a corresponding country and NACE code of the company were assigned to a company.

2.2 Banks

Our sample of banks consists of two groups. First, group-1 banks are selected from banks included in the 2011 EBA disclosures. The EBA disclosures are suitable for constructing sufficiently granular balance sheet structures. Furthermore, Bureau van Dijk's (BvD) Bankscope data on individual banks' balance sheet aggregates of total assets (TA_{*i*}), interbank borrowing and lending, customer loans, capital position and risk-weighted assets were used as a supplementary source to retrieve the main risk/return characteristics of the group-1 banks.

While the EBA disclosures encompass the largest EU banking groups, they cover only a fraction of the total EU banking sector. To improve coverage we apply a simulation approach to generate the second group of banks (group-2 banks). For that purpose, BvD Bankscope data on total assets of 500 banks in the countries covered by the group-1 banks was used. Table 2 on page 42 shows even bigger heterogeneity of banks' sizes compared to that of nonbank firms (in particular in France and UK). As the Bankscope data is less granular than the EBA disclosures used for the group-1 banks, we applied proportionality rules to assign missing balance sheet items for the group-2 banks. Specifically, for each group-2 bank j we approximated its interbank placements, interbank deposits, corporate lending and capital by taking the average corresponding category for banks covered by the 2011 EBA disclosures in a country of bank j and scaled this figure to the ratio of total assets of bank j

	(a) Coverage	
ltem	Description	Sources
Banks	Group 1: as identified in 2011 European Banking Authority (EBA) Disclosures; 80 banks from 21 EU countries ¹ ; group 2: BvD universe of 500 banks (beyond group 1)	EBA, Hałaj and Kok (2014), BvD Bankscope Bloomberg and own calculations I S EBA, BvD and own calculations f ECB and own calculations
Nonbank corporations	Members of the benchmark equity indices in the countries covered by EBA disclosures and Hałaj and Kok (2014); total 900 firms	Bloomberg and own calculations
	(b) Individual company level attributes	
ltem	Description	Sources
Banks	Total assets, interbank assets, securities, securities marked to market, equity, core tier 1 capital, interbank liabilities, country of incorporation	EBA, BvD and own calculations
Banks	Loans to nonfinancial corporations: calculated by applying the average country ratio of such loans to total assets based on the ECB Monetary Financial Institutions (MFI) balance sheet data set	EBA, Hałaj and Kok (2014), BvD Bankscope Bloomberg and own calculations I S EBA, BvD and own calculations f ECB and own calculations
Banks	CDS spreads of senior debt with 5-year maturity, and long-term issuer ratings by Moody's, Fitch and S&P	Bloomberg
Nonbank corporations	Total assets, total equity, total liabilities, NACE code, CDS spreads of senior debt with 5-year maturity, and long-term issuer ratings by Moody's, Fitch and S&P	Bloomberg
Nonbank corporations	Loans from banks: calculated by applying the average country ratio of loans to total assets of NFCs based on the ECB Euro Area Accounts data set	ECB and own calculations

TABLE 1	Description	of the data ir	puts. [Table	continues of	on next page.]
---------	-------------	----------------	--------------	--------------	----------------

¹ AT, BE, CY, DE, DK, ES, FI, FR, GB, GR, HU, IE, IT, LU, MT, NL, NO, PL, PT, SE and SI.

and average total assets of group-1 banks in the same country as group-2 bank j. Obviously, there are caveats to this approach. First and foremost, while the group-1 banks generally cover the largest and most predominant banks in their resident country

	(c) Lending relations and other supportive val	riables
ltem	Description	Sources
Lending relationship	Defined as the number of loans with different banks; average figures by country and NACE sector were applied based on the data provided through the National Central Banks	ECB calculations
Interest rates on loans by size and country	Average interest rates on loans by size of loan and by country based on the ESCB MIR data; categories of loans as follow: small (below 0.25 EUR mn), medium (equal or above 0.25-1 EUR mn), and large (over 1 EUR mn)	ECB and own calculations
Expected default frequencies	Average of expected default frequencies within 1 year by country and NACE	Moody's KMV and own calculations

TABL	.E 1	Continued.
		oonanaoa.

Source: own compilation.

they are not necessarily representative of the smaller (group-2) banks, especially in some of the larger EU countries with less concentrated banking sectors, such as Germany and Italy. This caveat notwithstanding, the proposed extension of the sample is important for better coverage of corporate lending portfolios, implying in turn a broader bank-firm lending network derived via the algorithm presented in Section 3.2.

In addition to the balance sheet information, the network formation algorithm described in more detail below also requires some market-based indicators to gauge counterparty risk. For this purpose, we apply bank-level CDS spreads and long-term issuer ratings. For those banks/firms where CDS spreads and issuer ratings were not available, we used sector peer group averages within countries. If no relevant peer group existed within the country, the EU average of the sectoral peer group was applied.

2.3 Lending relations

In the model, banks and nonbank companies are linked via their lending relationship. We consider two types of links, treated separately: (i) bank-to-nonbank and (ii) bankto-bank. The likelihood of the interbank linkages is modeled using the approach of Hałaj and Kok (2013) who define a probability map of linkages based on the 2011 EBA

TABLE 2 Data:	main s	itatistics	s per cc	ountry.	[Table (continu	i uo sər	next thi	ree pa	ges.]								
							וחאו (מ)			0								
	АТ	BE	DE	Ы	ES	Ē	FR	GB	GR	ш	F	R	N	Ы	SE	S	Other	Total
Banks	25	6	240	47	29	10	34	52	9	7	40	17	33	4	13	10	10	586
Nonbanks	18	19	28	147	27	131	37	97	57	43	35	24	23	46	26	2	126	891
Total	43	28	268	194	56	141	71	149	63	50	75	41	56	50	39	17	136	1477
						q)) Total ≨	assets	(€ billi	(suo								
	AT	BE	DE	A	ES	Ē	FR	GB	GR	ш	F	Ł	S	Ы	SE	S	Other	Total
Banks																		
25th percentile	0.6	5.2	1.7	0.3	9.4	6.7	6.6	4.3	32.2	3.0	4.6	6.1	0.8	68.8	9.0	0.9	2.9	1.8
Median	3.5	22.1	4.6	1.	29.4	24.8	13.4	14.5	63.1	33.1	12.8	11.3	3.1	78.9	34.5	1.6	17.2	6.2
75th percentile	17.5	214.2	9.6	5.8	65.4	31.5	68.2	62.0	80.1	73.0	40.5	76.5	9.2	85.6	187.4	3.7	35.9	24.7
SD	44.0	110.5	125.0	66.8	218.7	26.5	472.7	395.5	32.8	51.0	156.6	284.4	44.6	26.1	163.7	3.5	17.6	209.1
Nonbanks																		
25th percentile	2.0	3.6	21.1	0.0	4.4	0.0	21.3	4.0	0.1	0.0	4.8	5.0	1.4	0.1	5.0	1.3	0.0	0.1
Median	5.5	6.7	33.7	0.2	12.2	0.2	31.0	9.7	0.4	0.2	12.0	12.7	3.1	0.6	8.8	1.5	0.1	1.7
75th percentile	11.7	17.8	114.1	1.2	35.3	1.7	61.5	31.3	1.3	1.0	46.4	29.2	13.8	3.3	26.7	2.0	1.6	9.4
SD	10.9	52.8	142.3	58.6	33.4	61.1	129.0	104.3	2.7	7.0	81.4	88.2	23.9	15.4	12.9	0.8	24.4	70.6

Journal of Network Theory in Finance

www.risk.net/journal

Emergence of the EU corporate lending network 43

www.risk.net/journal

Journal of Network Theory in Finance

www.risk.net/journal

AT BE DE DK ES F1 FR GB GR IE NL NO PT SE SI Other Total Nonbanks Nonbanks 96 62 44 69 95 70 67 66 52 32 96 57 84 67 96 104 116 33 89 104 110 98 50 104 115 225 90 96 277 87 129 124 110 35 40 75 63 42 86 50 40 157 50 35 40 75 55 53 124 110 37 86 58 33 40 75 55 54 154 104 75 Banks 25 41 71 88 50 84 56 75 63 154 154 104 75 S5 102 104 </th <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>(i) Ave</th> <th>erage</th> <th>CDS (</th> <th>(sdq</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>								(i) Ave	erage	CDS ((sdq									
Nonbanks Nonbanks 25th percentie 96 62 44 69 95 84 59 70 67 96 52 32 96 67 84 67 96 75th percentie 96 102 55 87 96 104 77 84 96 104 102 96 127 87 104 102 96 127 87 104 102 98 95 104 115 225 90 96 277 87 104 102 98 95 104 115 225 90 96 277 87 104 102 167 104 71 188 50 84 96 179 179 86 58 241 59 154 104 104 104 104 104 104 104 104 104 104 104 104 104 104 104 104 104 10		AT	BE	DE	Ъ	ES	Ē	FR	GB	GR	⊔	⊨	Ł	0 N	Ч	SE	ទ	Other	Total	
25th percentile 96 62 44 69 58 84 56 7 96 57 84 67 69 75th percentile 110 133 89 104 140 110 98 95 104 155 90 96 57 84 67 96 75th percentile 110 133 89 104 140 110 98 65 104 165 85 29 35 40 75 Banks 25th percentile 102 167 104 71 188 50 84 96 67 193 179 86 58 215 154 124 104 75th percentile 102 167 104 71 188 50 84 678 103 179 86 58 2154 154 104 75th percentile 102 167 104 71 86 58 216 104	Nonbanks																			
Median 96 102 55 87 96 104 77 84 96 104 115 225 90 96 227 87 129 124 110 35 SD 22 42 48 32 41 61 60 29 197 50 75 63 42 85 24 124 110 SD 22 42 75 53 42 85 24 75 75 63 42 75 74 124 110 SD 102 167 104 71 188 50 84 96 75 83 73 75 7 7 75 7 7 75 7 7 75 7 7 75 7 7 7 75 7 7 7 7 7 7 7 7 7 7 7 7 7 7 104	25th percentile	96	62	44	69	95	84	59	20	67	67	96	52	32	96	57	84	67	69	
75th percentile 10 133 89 104 110 98 50 155 50 52 87 129 124 110 SD 22 42 48 32 41 61 60 29 197 50 75 63 42 85 40 75 Banks 25th percentile 102 167 104 71 188 50 84 98 671 193 179 86 58 154 154 104 75 SD 20 167 104 71 188 50 84 98 671 193 179 86 53 154 154 104 75th percentile 102 167 104 71 188 50 84 58 68 53 154 154 104 75th percentile 102 167 114 19 8 0 27 5 0	Median	96	102	55	87	96	104	11	84	96	104	166	83	58	151	64	104	102	96	
SD 22 42 48 32 41 61 60 29 197 50 75 63 42 85 40 75 Banks 25th percentile 102 167 104 71 188 50 84 98 67 193 179 86 58 218 59 154 104 75th percentile 102 167 104 71 188 50 84 98 673 179 86 58 214 59 154 104 75th percentile 102 167 104 71 188 50 84 200 179 86 58 214 59 154 104 75th percentile 102 104 71 188 50 84 200 179 86 58 214 59 154 104 75th percentile 102 104 1 2 10 14 19 </td <td>75th percentile</td> <td>110</td> <td>133</td> <td>89</td> <td>104</td> <td>140</td> <td>110</td> <td>98</td> <td>95</td> <td>104</td> <td>115</td> <td>225</td> <td>6</td> <td>96</td> <td>227</td> <td>87</td> <td>129</td> <td>124</td> <td>110</td> <td></td>	75th percentile	110	133	89	104	140	110	98	95	104	115	225	6	96	227	87	129	124	110	
Banks Banks Solution 102 167 104 71 188 50 84 98 677 193 179 86 58 216 154 154 164 75th percentile 102 167 104 71 188 50 84 98 678 193 179 86 58 241 59 154 104 104 75th percentile 102 167 104 71 188 50 84 98 678 109 86 58 241 59 154 154 104 75th percentile 102 167 104 71 188 50 84 200 179 86 58 241 59 154 104 75th percentile 102 167 11 12 11 2 10 14 19 8 20 16 104 75th bercentile 29 15 16	SD	22	42	48	32	41	61	60	29	197	50	75	63	42	85	29	35	40	75	
25th percentile 102 167 104 71 188 50 84 98 677 193 179 86 58 218 59 154 164 104 75th percentile 102 167 104 71 188 50 84 98 678 193 179 86 58 241 59 154 104 75th percentile 102 167 104 71 188 50 84 98 678 193 179 86 58 241 59 154 104 75th percentile 102 28 0 1 2 10 14 19 8 0 27 5 0 16 69 75th 20 28 61 74 19 8 0 27 5 0 16 69 75th 8 7 8 7 19 8 0 27 <	Banks																			
Median 102 167 104 71 188 50 84 98 678 193 179 86 58 232 59 154 154 104 75th percentile 102 167 104 71 188 50 84 200 179 86 58 241 59 154 164 104 SD 20 28 6 7 10 14 19 86 58 241 59 154 164 104 SD 20 28 6 1 2 10 14 19 86 5 164 164 104 SD 20 2 3 0 1 2 10 14 19 86 5 164 164 104 SD 2 2 3 16 69 5 16 16 69 164 164 16 16 16	25th percentile	102	167	104	7	188	50	84	98	677	193	179	86	58	218	59	154	154	86	
75th percentile 102 167 104 71 188 50 84 96 58 241 59 154 104 104 SD 20 2 37 0 1 2 10 14 19 86 58 241 59 154 104 69 SD 20 28 6 7 10 7 10 7 5 0 16 69 AT BE DE DK ES FI FR GB T NI NO PT S 0 16 69 AT BE DF DK ES FI FR GB T NI NO PT S 0 16 69 AT BE DF DK ES FI FK GB T NI NO PT S 0 16 69 30 AT BE </td <td>Median</td> <td>102</td> <td>167</td> <td>104</td> <td>7</td> <td>188</td> <td>50</td> <td>84</td> <td>98</td> <td>678</td> <td>193</td> <td>179</td> <td>86</td> <td>58</td> <td>232</td> <td>59</td> <td>154</td> <td>154</td> <td>104</td> <td></td>	Median	102	167	104	7	188	50	84	98	678	193	179	86	58	232	59	154	154	104	
SD 20 28 6 2 37 0 1 2 10 14 19 8 0 27 5 0 16 69 AT BE DE DK ES FI F C 10 14 19 8 0 27 5 0 16 69 AT BE DF DK ES F C B IF NI NI NI P P P 10 25th percentile 2.91 3.00 3.06 3.00 3.00 3.00 2.01 3.00 3.00 3.00 Median 3.12 3.24 3.36 3.36 3.21 3.06 3.23 3.30 3.23 3.30 3.00 2.34 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.36 3.	75th percentile	102	167	104	7	188	50	84	98	684	200	179	86	58	241	59	154	154	104	
(j) Lending relationship: nonbanks AT BE DK ES FI FI N Other Total 25th percentile 2.91 3.06 3.00 3.00 3.00 3.00 2.04 3.00 <td>SD</td> <td>20</td> <td>28</td> <td>9</td> <td>2</td> <td>37</td> <td>0</td> <td>-</td> <td>0</td> <td>10</td> <td>14</td> <td>19</td> <td>8</td> <td>0</td> <td>27</td> <td>2</td> <td>0</td> <td>16</td> <td>69</td> <td></td>	SD	20	28	9	2	37	0	-	0	10	14	19	8	0	27	2	0	16	69	
AI BE DF DK FS FI FR GB FI NL NO PT SE SI Other Total 25th percentile 2.91 3.00 3.06 3.57 3.00 3.00 3.00 2.10 SO 3.00 <																				
AT BE DF DK ES FI FR GB GF IT NL NO PT SE SI Total 25th percentile 2.91 3.00 3.06 3.57 3.00 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td>n) re</td><td>hunu</td><td>ופומוור</td><td>unsing.</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>							n) re	hunu	ופומוור	unsing.										
25th percentile 2.91 3.00 3.06 3.00		АТ	BE	DE	Ŋ	ES	Ē	FR	GB	GR	ш	╘	R	0 N	Ы	SE	เง	Other	Total	
Median 3.12 3.24 3.24 3.36 3.12 3.06 3.24 3.06 3.05 3.01 3.12 3.24 3.24 3.12 75th percentile 3.45 3.30 3.12 3.36 4.02 3.36	25th percentile	2.91	3.00	3.06	3.06	3.57	3.00	3.00	3.00	3.00	3.00	2.94	3.00	3.00	2.10	3.00	3.06	3.00	3.00	
75th percentile 3.45 3.30 3.12 3.36 4.02 3.36 3.36 3.36 3.36 3.30 3.51 3.36 3.12 2.58 3.36 3.30 3.36 3.36 3.36 S.36 S.36 S.36 S.36 S.36	Median	3.12	3.24	3.12	3.24	3.90	3.24	3.36	3.12	3.06	3.06	3.21	3.12	3.00	2.34	3.36	3.24	3.24	3.12	
SD 0.39 0.18 0.30 0.21 0.48 0.18 0.21 0.21 0.21 0.18 0.27 0.15 0.12 0.27 0.21 0.15 0.21 0.30	75th percentile	3.45	3.30	3.12	3.36	4.02	3.36	3.36	3.36	3.36	3.30	3.51	3.36	3.12	2.58	3.36	3.30	3.36	3.36	
	SD	0.39	0.18	0.30	0.21	0.48	0.18	0.21	0.21	0.21	0.18	0.27	0.15	0.12	0.27	0.21	0.15	0.21	0.30	

Emergence of the EU corporate lending network 45

www.risk.net/journal

		Total	1443		Total	0.4	3.5	14.2	143.0		Total	0.0	0.5	2.6	18.0
		Diversified	7		Diversified	0.4	0.9	4.0	6.3		Diversified	0.1	0.3	1.2	2.0
		Utilities	32		Utilities	7.9	15.6	49.3	69.1		Utilities	2.5	5.7	15.8	17.1
		Tech.	42		Tech.	0.0	0.1	1.9	4.5		Tech.	0.0	0.0	0.5	1.3
		Energy	54		Energy	1.1	4.1	14.6	73.3		Energy	0.2	1.1	4.6	20.5
	ities	Basic materials	72	(suc	Basic materials	0.4	2.7	13.6	26.7	(s billions)	Basic materials	0.1	0.9	3.9	7.3
	per of ent	Comm.	78	ts (€ billic	Comm.	0.1	1.9	13.5	29.7	banks (€	Comm.	0.0	0.6	3.5	8.9
	a) Total numt	Consumer, cyclical	119	b) Total asse	Consumer, cyclical	0.0	1.0	6.9	35.2	rrowing from	Consumer, cyclical	0.0	0.3	1.5	10.2
)	Consumer, noncyclical	151		Consumer, noncyclical	0.1	1.2	6.9	16.4	(c) Bo	Consumer, noncyclical	0.0	0.3	2.1	4.4
s per sector.		Industrial	172		Industrial	0.1	1.3	5.4	15.4		Industrial	0.0	0.4	1.7	4.1
nain statistic		Financial	716		Financial	1.3	5.5	24.6	196.4		Financial*	0.0	0.5	4.2	37.1
TABLE 3 Data: n						25th percentile	Median	75th percentile	SD			25th percentile	Median	75th percentile	SD

www.risk.net/journal

46 G. Hałaj et al

			(d) Inte	rest rate on c	corporate	loans (%)					
	Financial*	Industrial	Consumer, noncyclical	Consumer, cyclical	Comm.	Basic materials	Energy	Tech.	Utilities	Diversified	Total
25th percentile	1.26	2.13	2.28	2.13	2.01	2.13	2.40	2.04	2.34	1.89	2.04
Median	1.92	2.19	2.73	2.61	2.25	2.73	2.58	2.31	2.88	2.13	2.46
75th percentile	3.45	3.18	3.51	4.77	3.48	3.48	4.35	2.94	3.69	3.78	3.72
SD	2.28	1.17	1.29	1.50	1.71	1.38	1.32	1.32	1.41	1.68	1.59
				(e) Average	CDS (bp	s)					
	Financial	Industrial	Consumer, noncyclical	Consumer, cyclical	Comm.	Basic materials	Energy	Tech.	Utilities	Diversified	Total
25th percentile	84	96	67	104	64	111	32	84	82	115	71
Median	104	96	67	104	100	151	67	84	83	115	98
75th percentile	110	110	80	104	103	194	67	84	95	115	104
SD	85	30	19	49	73	59	28	0	25	0	70

TABLE 3 Continued.

www.risk.net/journal

Journal of Network Theory in Finance

Emergence of the EU corporate lending network 47

	Financial*	Industrial	Consumer, noncyclical	Consumer, cyclical	Comm.	Basic materials	Energy	Tech.	Utilities	Diversified	Total
25th percentile	28	29	28	28	28	27	24	29	26	30	28
Median	29	31	31	29	32	32	29	33	32	31	31
75th percentile	34	34	34	34	34	34	35	34	36	33	34
SD	9	9	9	7	7	7	13	ი	7	2	7
			(g) Lenain	g relation (av	erage ruu		(2				
	Financial*	Industrial	Consumer, noncyclical	Consumer, cvclical	Comm	Basic materials	Fnerav	Tech	Utilities	Diversified	Total
25th percentile	2.90	3.12	3.06	2.58	3.00	3.00	3.00	3.00	2.91	2.88	3.00
Median	3.24	3.36	3.36	3.06	3.00	3.12	3.00	3.00	2.91	3.06	3.12
75th percentile	3.24	3.36	3.36	3.36	3.15	3.36	3.00	3.36	2.91	3.24	3.36
SD	0.24	0.24	0.30	0.36	0.36	0.27	0.18	0.33	0.15	0.33	0.30

TABLE 3 Continued.

Journal of Network Theory in Finance

www.risk.net/journal

48 G. Hałaj et al



FIGURE 1 Total assets of companies selected to the model (2012; € millions).

disclosures of country and product breakdowns of banks' exposures. For the bank-firm interlinkages, in the absence of concrete data of individual bank and firm lending relationships, we necessarily have to make a number of assumptions. To estimate volumes of banks' corporate lending portfolios we use the country-specific ratios derived from the ESCB Monetary Financial Institutions (MFI) consolidated balance sheet data set. For the non-EU countries we use congruent, national banking statistics. Based on the MFI data, the country-specific ratios of the loans to the nonfinancial sector to the total balance sheet of the MFIs are calculated and applied to the individual banks' balance sheet figures in order to obtain the proxy for the bank-specific corporate loan portfolio.

On the nonbank firm side of the lending relationship, an approximation of the share of firms' external financing that derives from bank borrowing can be made using country-specific indicators based on Euro Area Accounts (EAA), or flow of funds, statistics comprising the full set of financial assets and liabilities of all resident

www.risk.net/journal

Source: Bloomberg and own calculations.

FIGURE 2 Loans from banks (2012).



Source: Bloomberg, EA accounts and own calculations.

institutional sectors including the nonfinancial and financial corporations. Based on these statistics, country aggregate ratios of the nonfinancial corporations' bank borrowing to total assets within a given country are calculated and then applied uniformly to the total assets of individual firms in our sample. The outcomes of these approximations are presented in Figure 2 and Figure 3 on the facing page. The charts show the individual banks' corporate loan portfolio and the individual firms' loans from banks.

Information on the loans was enriched with data on lending rates, which is another measure of the credit risk of individual firms. The average interest rates on loans by size of loan and by country were obtained from the MFI interest rate statistics of the European Central Bank (ECB), including breakdowns of the interest on small loans (below ≤ 0.25 million), medium-sized loans (equal or above $\leq 0.25-1.00$ million) and large-sized loans (above ≤ 1 million).

The mechanism to reconstruct the connection between banks and firms is described in Section 3. National credit register data, which includes anonymous (and





Source: Bloomberg, MFI and own calculations.

confidential) individual loans data with geographical and NACE sector code breakdowns covering most euro area countries, was a key additional source of information for the construction of the bank–firm network of lending relationships. This allows for estimating the average number of loans that individual firms have with different banks. This was a crucial component in determining the probability that firms and banks are linked to each other, as the average number of bank–firm interlinkages gleaned from the credit register data helps inform the range of banks that each firm is likely to borrow from (via the bank–firm probability map). Table 2 on page 42 and Table 3 on page 46 show an average number of three linkages of nonbank firms with banks, a number which is steady across countries and sectors.

3 MODEL

In this section, we describe the assumptions behind the corporate network formation process followed by a specification of an algorithm used to generate the network of loans that banks extend to other banks and nonbank firms. Finally, we adopt the

www.risk.net/journal

traditional cascade approach to study potential contagion effects of corporate defaults transmitted via the generated networks of loan exposures.

3.1 Corporate network formation

The model corporate network formation describes how a system of banks and nonbank firms linked by exposures related to lending activity of banks emerges.³

The simulation of bank-firm networks follows an extension of Hałaj and Kok (2014) whereby the linkages of exposures are assigned sequentially between nodes in a given predefined set. However, the introduction of nonbank firm nodes requires some important modifications and extensions to capture the complexity of the bankfirm network. First, the interbank linkages and links between banks and nonbank firms are created independently. When choosing a set of counterparties to optimize the structure of the lending portfolio, banks pick counterparties first for banks and later for nonbank corporates. Second, nonbank firms restrict the preferred number of banks that provide financing to them. In reality, the lending relationships are likely to be formed by nonbank firms first deciding whether they need external financing and from which banks they want to apply for loans. Only in the second step will banks then decide who to lend to and on what terms. Finally, primarily reflecting liquidity management purposes once the bank-firm relationships are formed, banks will decide on interbank lending and borrowing. Technically, however, the outcomes of the algorithm presented below are invariant to the order in which interbank or nonfinancial corporate portfolios are optimized.

The investment choice of banks is assumed to be derived from an optimization of the return from loans (expected return), the risk of interest payments and default (volatility of returns) and diversification (correlation of returns from loans to extended to different counterparties). The contractual loan rate is proxied using official ECB statistics; namely, data on MFI interest rates (MIR) by country and by loan size. The country-specific lending rates are in turn combined with the firms' credit ratings to derive firm-specific rates. Specifically, the ratings are translated into the credit risk spreads and for each bank the deviation of its spread from the average in the sample is attached to the country-specific rate yielding the lending rate on loans to that firm. The risk of return on a loan extended to a company i is measured as

$$\sigma_i = \sqrt{(1 - p_i + (1 - \lambda)^2 p_i)(\bar{r}_i^2 + \sigma_{r,i}^2) - (1 - p_i + (1 - \lambda)p_i)^2 \bar{r}_i^2}, \quad (3.1)$$

³ Notation: $\bar{K} = \{1, 2, ..., K\}$ for each natural number K; $\mathcal{M}^{m \times n}$ is the set of real values matrixes with *m* rows and *n* columns; \mathcal{M}^n is a square matrix of size *n*; $A \times B$ – for *N*-dimensional vectors *A* and *B* producing a vector $[A_1B_1, A_2B_2, ..., A_NB_N]$. For $z \in \mathbb{R}$, $\lceil z \rceil$ is the smallest number in \mathbb{Z} which is higher than or equal to *z*.

where

- \bar{r}_i is the lending rate to a company i,
- $\sigma_{r,i}$ is the volatility of the lending rate to *i*,
- *p_i* is default probability of *i*, equivalent to the estimated credit risk spread (*p_i* := {spread}_i/{lgd}, for lgd = 0.5),
- λ is the loss given default parameter.

In fact, $(\sigma_i)^2$ is the variance of a random variable X_i combining the interest rate risk (assumed to be a normally distributed random variable r_i with mean \bar{r}_i and standard deviation $\sigma_{r,i}$) and the default risk (assuming the firm defaults on its interest payments with probability p_i):

$$X_i = \begin{cases} r_i & \text{with probability } 1 - p_i, \\ (1 - \lambda)r_i & \text{with probability } p_i. \end{cases}$$

The correlation structure of returns from corporate portfolios can be either estimated or set via some rules of thumb. In the absence of a reliable proxy the currently applied rule of thumb assumes zero correlation.

Individual, optimization-based decisions of banks and nonbank firms to invest and finance their activities contribute to an aggregate network of exposures between agents but within an assumed structure of lending relationship in the market. The lending relationship structure is derived from a probability map defining the likelihood that a given bank extends a loan to another firm – a bank or a nonbank company. For interbank lending, the probability map from Hałaj and Kok (2013) is used for the likelihood of linkages between banks in the EBA disclosure sample. The estimates of probabilities are based on the banks' balance sheet data and in particular on geographical breakdowns of loans provided in the EBA institution-specific disclosures of EU banks. For the links between the banks from the EBA sample and their non-EBA sample domestic counterparties, since we do not have any data for the estimations, we assign a stylized probability of 1%. Finally, we conservatively assume 0% probability of linkages between banks outside the EBA disclosures sample. The endogenous interbank relationships that are formed using our optimization algorithm will thus be a subset of the preselected group of banks that individual banks, according to the probability map, are likely to be related to. For the nonfinancial companies in the sample we use Credit Register data covering most euro area countries, and which contains a geographical breakdown of loans. We calculate the average geographical structure of loans for broad NACE sectors in a country. For those corporate agents in the model where no geographical structure is available, we assume some stylized

www.risk.net/journal

probabilities of linkages (domestic is 90% and foreign is 1%).⁴ The combined interbank and nonfinancial corporate map would create a large map that is used to draw pairs of agents potentially entering a lending contract.

A network of interbank exposures and corporate lending is generated under the assumption of optimizing the behavior of banks and nonfinancial firms. Moreover, for what concerns the banks' portfolio optimization we assume that the interbank and corporate loan portfolios are optimized separately. This assumption tries to reflect the different risk management treatment of these two types of portfolios, whereas interbank lending is closer to liquidity management practices and corporate lending is part of the investment portfolio. These important differences do not exclude the expected returns and related risks as important drivers of the allocation of investments in each of the two portfolios. We take the volumes of the liquidity and investment portfolios as results of an asset and liability management (ALM) process. Banks specify the amount of liquidity that they need to acquire from the interbank market and the pool that they can offer to lend to other banks. The separation of the two investment subportfolios (lending to customers and interbank lending) can stem from a strategic ALM allocation which may be expected to remain relatively stable over time (Adam 2008). Therefore, we expect that banks first choose the target for volumes of nonfinancial corporate and interbank portfolios, respectively, and then optimize within these two subportfolios. On the asset side, their preferred structure of investment maximizes the risk-adjusted return. On the liability side, the preferred funding structure is selected to minimize the funding risk. Funding risk is associated with the stability of the funding relationship. It is measured by a matrix $D \in \mathcal{M}^N$ which is a covariance matrix of random variables $Z_i, i \in \{1, \ldots, N\}$, corresponding to banks, taking a value $1 - \lambda$ with probability p_i and a value 1 with probability $1 - p_i$. Both the adjustment of returns by their risk, which is taken into account in the choice of the asset structure, and the funding risk for the selection of the funding structure introduce quadratic terms to the optimization programme and reflect the risk averseness of all the agents in the model.

At this stage of the algorithm we make an important deviation from Hałaj and Kok (2014) concerning funding relationships of nonbank corporations. In their financing decisions, most nonbank companies usually operate with only a very limited number of banks from which they borrow money to finance their projects (see, for example, Petersen and Rajan 1994). The Central Credit Registers data provides us with information about the number of banks that each company in the sample has in its funding portfolio. We use the estimated mean ($\mu_{c,n}$) and variance ($\sigma_{c,n}^2$) of

⁴ This approach is similar to, for example, the BIS report on OTC derivatives which also use some stylized probabilities where no information is available for the estimation of linkages (Bank for International Settlements 2013).

the numbers per country (*c*) and broad NACE sector (*n*) combined with the probability map to draw subsamples of bank lenders for each company. More precisely, for each nonfinancial corporation *i* we first use the probability map to preselect the group of banks G_i which may establish a lending relationship with that company. Second, from a lognormal distribution with mean $\exp(\mu_{c,n} + \sigma_{c,n}^2/2)$ and variance $\exp(2\mu_{c,n} + \sigma_{c,n}^2)/(\exp(\sigma_{c,n}^2) - 1)$ we draw the ultimate number of banks N_i and randomly pick $\lceil N_i \rceil$ banks from G_i .⁵ Conversely, it is assumed for banks that the degree constraint in their interbank funding sources is very lax meaning that banks may form relatively large number of lending relationships on the interbank market. Subsequently, after these two steps the optimization of funding sources is performed, as in Hałaj and Kok (2014).

The network formation algorithm is ultimately run to match in an iterative way the corporate lending portfolios of banks to the funding preferences of other banks and the external financing needs of nonbank firms. The algorithm is analogous to the one proposed by Hałaj and Kok (2014). In each iteration, or "step" of the algorithm, banks and firms choose their preferred counterparties and volume of loans that they are most willing to grant or accept. The inherent mismatch between supply side (related to extended loans) and demand side of the market (related to funding sought by banks and nonbank firms) is assumed to be leveled out by a bargaining game played by pairs of agents.⁶ The aggregate outcomes of the game may still leave a mismatch. In such a (likely) case the next iteration (step) is activated to allow agents to enlarge their sets of potential counterparties and to reoptimize the preferred structure of loans. The steps are repeated until convergence. We present the procedure in a rigorous, algorithmic way in Section 3.2. A snapshot of the outcome of the network formation process is presented in Figure 4 on page 64.

3.2 Algorithm

The algorithm consists of a sequence of the three rounds preceded by an initialization phase (points (1)–(9) of the algorithm specified in this section). The rounds reflect the three main building blocks of the process describing the emergence of the lending network among banks and nonbank firms. The first round (point (10) of the algorithm) implements the optimization of banks' loan portfolios. The second round describes

www.risk.net/journal

⁵ The lognormal distribution is generated by a normally distributed variable with mean $\mu_{c,n}$ and variance $\sigma_{c,n}^2$. N_i is obtained by drawing a number from the lognormal distribution and taking its integer part.

⁶ The bilateral games are myopic in the sense that they partly ignore the links already established with other agents. It is only a partial omission since each agent defines the set of their potential counterparties at the beginning of the formation process and then at each iteration they randomly enlarge it. It is a different approach than the one taken by Manea (2014) where the number of already established connections matter for the choice of the new linkages.

the optimization of the funding sources, both of banks and nonbank firms (point (11)). The third round covers the bargaining game (points (12)-(13)). The three rounds are repeated sequentially to allocate loans in banks' interbank and corporate portfolios to interbank and corporate funding portfolios of other banks and nonbank firms. The accruing of loans in the emerging network is performed in point (14) of the algorithm.

- (1) Define the nodes, ie, the two sets of tuples representing banks (*b*[*i*]) and nonbank firms (*c*[*i*]):
 - Set **B**:

$$b[i] = (i, a_i^{\mathrm{T}}, a_i^{\mathrm{B}}, a_i^{\mathrm{C}}, l_i^{\mathrm{B}}, e_i, e_i^{\mathrm{B}}, e_i^{\mathrm{C}}, r_i, \sigma_i^{\mathrm{B}} := [\sigma_1, \dots, \sigma_N])$$

and $\#\mathcal{B} = N$

• Set C:

$$c[i] = (N + i, a_{N+i}^{T}, l_{i}^{C}, e_{N+i}, r_{N+i}, \sigma_{i}^{C} := [\sigma_{N+1}, \dots, \sigma_{N+M}])$$

and #C = M

and
$$\#C \equiv$$

where

- a^{T} total assets
- $a^{\rm B}$ interbank assets of banks
- a^{C} volume of corporate (nonbank related) lending portfolio of banks
- $l^{\rm B}$ interbank liabilities of banks
- $l^{\rm C}$ liabilities of nonbank firms related to borrowing from banks
- e (for banks) total regulatory capital; (for nonbank firms) total capital
- e^{B} banks' capital allocated to interbank lending
- e^{C} banks' capital allocated to corporate lending to nonbank firms
- *r* interest rate on loans, specific to a country and (NACE) sector of a borrower
- $\omega^{\rm B}$ risk weights of the interbank exposures
- $\omega^{\rm C}$ risk weights of the nonbank corporate loans
- σ^{B} , σ^{C} volatility of lending rates computed based on reference interest rates and spreads (or PDs) related to country/sector specific credit risk (see formula (3.1))
- (2) Define probability maps describing the likelihood that a pair of nodes is connected by a lending exposure. This consists of:

- a square matrix $P^{\text{geo},B} \in \mathcal{M}^N$ describing the likelihood that banks connect to each other;
- a matrix $P^{\text{geo},C} \in \mathcal{M}^{M \times N}$ describing the likelihood that banks connect to nonbank firms; the shape of the matrix reflects an assumption that nonbank firms do not lend to banks.
- (3) Define a correlation structure of returns from loans $Q^{B} \in \mathcal{M}^{N}$ to banks and $Q^{C} \in \mathcal{M}^{M}$ to the nonbank corporate sector; $D \in \mathcal{M}^{N}$ is the funding diversification matrix (common for banks and nonbank firms since funding in the model is only provided by the *N* banks).
- (4) Initialize the exposure matrixes:

$$L^{\mathrm{B},0} \in \mathcal{M}^{N}, \qquad \forall (i,j) \in \bar{N} \times \bar{N}L^{\mathrm{B},0}_{ij} := 0,$$
$$L^{\mathrm{B},\mathrm{C},0} \in \mathcal{M}^{M \times N}, \quad \forall (i,j) \in \bar{M} \times \bar{N}L^{\mathrm{B},\mathrm{C},0}_{ij} = 0.$$

- (5) Initialize vectors of unallocated assets $a^{B,1} := a^B, a^{C,1} := a^C$ and liabilities $l^{B,1} := l^B$ and $l^{C,1} := l^C$. Initialize the vectors of unallocated capital $e^{B,1} = e^B$ and $e^{C,1} = e^C$. The entries of the vectors are gradually reduced in each step of the algorithm as banks' assets are matched with the assets and liabilities of banks and nonbank firms creating a network of direct exposures.
- (6) Initialize the counterparty sets for all banks and firms:
 - \$\bar{B}_i^{B,A,0} = \emptyset\$ is a set of preferred interbank debtors of bank *i* (ie, *i* is willing to place deposits in banks from \$\bar{B}_i^{B,A,0}\$);
 - $\bar{B}_i^{B,F,0} = \emptyset$ is a preferred set of interbank creditors of bank *i*;
 - $\bar{B}_i^{C,0} = \emptyset$ is the set of nonbank corporate firms to which *i* is willing to grant loans;
 - $\bar{C}_{j}^{B,0} = \emptyset$ is the set of banks from which a given nonbank firm is willing to borrow money to finance its activities.
- (7) Define an out-degree distribution of all nonbank corporate nodes: let *c* be a set of countries covering all nonbank firms in the sample and *n* be a set of NACE sectors. Let *c* be a mapping from nonbank indexes N + 1, N + 2, ..., N + M to *c* and let *n* be a mapping from nonbank indexes N + 1, N + 2, ..., N + M to *n*. Then, *m̃_j* be drawn from lognormal distribution created by exponential transformation of a normal distribution with mean μ_{c(j),n(j)} and σ_{c(j),n(j)}. Ultimately, the maximal number m_j of links incoming to c[j] ∈ C is given by [*m̃_j*].

- 58 G. Hałaj et al
 - (8) Set a precision $\epsilon > 0$ for the allocation of loan portfolios to the funding needs of the nonbank firms.
 - (9) Define bank-specific risk aversion parameters: κ_i for investment portfolio choice and κ_i^F for funding diversification decisions.
 - (10) [round 1, step k] For each $b[i] \in \mathcal{B}$ draw N numbers $(x_1^k, x_2^k, \dots, x_N^k)$ from a uniform distribution on [0, 1]

Let $B_i^{B,k} = \{j \mid x_j^K \leq P_{ji}^{\text{geo},B}\}$ be the set of potential new accepted interbank counterparties at step k. Then

$$\bar{B}_i^{\mathrm{B},A,k} = \bar{B}_i^{\mathrm{B},A,k-1} \cup B_i^{\mathrm{B},A,k}$$

is the new set of counterparties on the interbank market

For $\bar{B}_i^{C,k}$ the recursive procedure is similar. Let $B_i^{C,k} = \{j \mid x_j^k \leq P_{ji}^{\text{geo},B}\}$ be the set of potential new accepted nonbank corporate counterparties of bank *i* at step *k*. Then

$$\bar{B}_i^{\mathrm{C},k} = \bar{B}_i^{\mathrm{C},k-1} \cup B_i^{\mathrm{C},k}$$

is the new set of counterparties on the corporate lending market.

On the interbank market, banks solve the optimization problem, whereby for $i \in \overline{N}$:

they find a vector $[L_{1i}^k, \ldots, L_{Ni}^k] \in \mathbb{R}^N$ that maximizes a functional

$$J(L_{1i}^k,\ldots,L_{Ni}^k) = \sum_{j\in\bar{B}_i^{\mathrm{B},k}} r_j^k L_{ji}^k - \kappa_i (\sigma^{\mathrm{B}} L_{\cdot,i}^k)^{\mathrm{T}} \mathcal{Q}^{\mathrm{B}} (\sigma^{\mathrm{B}} L_{\cdot,i}^k), \qquad (3.2)$$

such that

$$\forall j \in \bar{N} \ L_{ji}^{k} \ge 0$$
$$\sum_{j=1}^{N} L_{ji}^{k} \le a_{i}^{B,k}$$
$$\sum_{j=1}^{N} \omega_{j}^{B} L_{ji}^{k} \le e_{i}^{B,k}$$

whereby $j \notin \bar{B}_i^{\mathrm{B},k} \Rightarrow L_{ji}^k = 0.$

Journal of Network Theory in Finance

www.risk.net/journal

On the nonbank corporate loan market, banks solve the following optimization problem, whereby for $j \in \{1, 2, ..., M\}$:

they find a vector $[L_{1j}^k, \ldots, L_{Mj}^k] \in \mathbb{R}^M$ that maximizes a functional

$$J(L_{1j}^k,\ldots,L_{Mj}^k) = \sum_{i\in\bar{B}_j^{\mathrm{C},k}} r_{N+i}^k L_{ij}^k - \kappa_j (\sigma^{\mathrm{C}} L_{\cdot,j}^k)^{\mathrm{T}} Q^{\mathrm{C}} (\sigma^{\mathrm{C}} L_{\cdot,j}^k), \quad (3.3)$$

such that

$$\forall i \in \bar{M} \ L_{ij}^k \ge 0$$

$$\sum_{i=1}^M L_{ij}^k \le a_j^{C,k}$$

$$\sum_{i=1}^M \omega_i^C L_{ij}^k \le e_j^{C,k}$$

whereby $i \notin \bar{B}_j^{C,k} \Rightarrow L_{ij}^k = 0.$

Let $L^{B,A,k}$ be a vector of some preferred allocation of placements on the interbank market (maximizing (3.2)) and let $L^{B,C,A,k}$ be a vector of preferred investment on the nonbank corporate lending market (maximizing (3.3)) after round 1 of step k.

(11) [round 2, step k] Banks and nonbank corporate firms decide about their preferred structure of funding from banks. They minimize the risk of obtaining funding/financing from banks that have the highest default risk which may break the established lending relationship. Moreover, firms take into account the number of banks that provide funding to individual firms (out-degree distribution).

For what concerns banks' funding for each $b[i] \in \mathcal{B}$ we draw N numbers x_j^k from a uniform distribution on [0, 1]. Then

$$B_i^{\mathrm{B},F,k} = \{j \mid x_j^k \leqslant P_{ij}^{\mathrm{geo},B}\}\$$

is a set of new potential interbank creditors of *i* and $\bar{B}_i^{B,F,k} = \bar{B}_i^{B,F,k-1} \cup B_i^{B,L,k}$ is a set of potential counterparties of *i* in step *k*.

With respect to nonbank firms' optimized borrowing from banks for each $c[i] \in C$ we draw N numbers x_j^k from a uniform distribution on [0, 1]. Let $\tilde{C}_i^{B,k} = \{j \mid x_j^k \leq P_{ij}^{geo,C}\}$ and let $C_j^{B,k}$ be a random subset of $\tilde{C}_i^{B,k}$, having at most m_i elements, which is the set of potential new accepted interbank counterparties at step k. The random sampling of the subsets $C_i^{B,k}$ is the following:

www.risk.net/journal

- if $\#\tilde{C}_i^{\mathrm{B},k} \leq m_i$ then $C_i^{\mathrm{B},k} := \tilde{C}_i^{\mathrm{B},k}$,
- if $\#\tilde{C}_i^{B,k} > m_i$ then we define a probability space on a family of m_i -element subsets of $\tilde{C}_i^{B,k}$ assigning to each element an equal probability

$$\frac{(\#\tilde{C}_{i}^{\mathrm{B},k}-m_{i})!(m_{i})!}{(\#\tilde{C}_{i}^{\mathrm{B},k})!}$$

and one subset $C_i^{B,k}$ is drawn with this probability.

Then

$$\bar{C}^{\mathrm{B},k}_i = \bar{C}^{\mathrm{B},k-1}_i \cup C^{\mathrm{B},k}_i$$

is a new set of counterparties on the nonbank corporate loan market.

In terms of optimization the following is assumed:

For $i \in \{1, 2, ..., N\}$ find a vector $[L_{i_1}^k, ..., L_{i_N}^k] \in \mathbb{R}^N$ that minimizes

$$F^{B}(L_{i1}^{k}, \dots, L_{iN}^{k}) = \kappa_{i}^{F}[L_{i1}^{k} \dots L_{iN}^{k}]D[L_{i1}^{k} \dots L_{iN}^{k}]^{T}$$
(3.4)

such that $j \notin \bar{B}_i^{B,F,k} \Rightarrow L_{ij}^k = 0$ and $\sum_{j=1}^N L_{ij}^k \leq l_i^{B,k}$. For $j \in 1, 2, ..., N$ find a vector $[L_{j1}^k, ..., L_{jN}^k] \in \mathbb{R}^N$ that minimizes

$$F^{C}(L_{j1}^{k}, \dots, L_{jN}^{k}) = \kappa_{j}^{F}[L_{j1}^{k}, \dots, L_{jN}^{k}]D[L_{j1}^{k}, \dots, L_{jN}^{k}]^{T}$$
(3.5)

such that $i \notin \bar{C}_j^{B,k} \Rightarrow L_{ji}^k = 0$ and $\sum_{i=1}^N L_{ji}^k \leq l_j^{C,k}$.

Let us denote a vector of the optimal structure of the preferred funding sources of bank *i* as $L_i^{B,F,k}$ (maximizer of the functional (3.4)) and let $L_i^{C,F,k}$ be a vector of the optimal structure of the preferred funding sources of a nonbank (maximizer of the functional (3.5)) after round 2 of step *k*.

(12) [round 3, step k] The negotiation round of the game is designed to reduce the results of divergence between preferences on the asset and liability side of agents' balance sheets. The outcome of the bargaining game is an agreed tentative volume of a loan granted by bank j to bank i (if $i \le N$) or a nonbank (if i > N). "Tentative" means that it may be subject to adjustment if the aggregate volume of loans exceeds the predefined one.

Four different outcomes are conceivable:

Case
$$L_{ij}^{B,A,k} > L_{ij}^{B,F,k}$$
:
maximize

$$G_{ij}^{k}(x) = [U_{ij}^{l,k*} - s_{ij}^{l,k}(x - L_{ij}^{\mathrm{B},F,k})][U_{ij}^{a,k*} - s_{ij}^{a,k}(L_{ij}^{\mathrm{B},A,k} - x)]$$
(3.6)

Journal of Network Theory in Finance

www.risk.net/journal

over $x \in [L_{ij}^{B,F,k}, L_{ij}^{B,A,k}]$, where $U_{ij}^{l,k*}$ and $U_{ij}^{a,k*}$ are utilities of the lending volumes optimal from borrower and lender perspective respectively, and $s_{ij}^{a,k}$ and $s_{ij}^{l,k}$ are measures of how much bank *i* and *j* are willing to deviate from their optimal investment and funding strategy, respectively. The sensitivity is inverse proportionate to the asset size, ie, $s_{ij}^{l,k} = 1/a_i^t$ and $s_{ij}^{a,k} = 1/a_j^t$.

Case
$$L_{ii}^{B,A,k} < L_{ii}^{B,F,k}$$
: analogously

The maximizer of the formula (3.6) is denoted $\bar{L}_{ij}^{B,k}$, which for all *i*s and *j*s creates a matrix of exposures $\bar{L}^{B,k}$.

Case
$$L_{ij}^{B,C,A,k} > L_{ij}^{C,F,k}$$
:
maximize

maximize

$$G_{ij}^{k}(x) = [U_{ij}^{l,k*} - s_{ij}^{l,k}(x - L_{ij}^{C,F,k})][U_{ij}^{a,k*} - s_{ij}^{a,k}(L_{ij}^{B,C,A,k} - x)]$$
(3.7)

over $x \in [L_{ij}^{C,F,k}, L_{ij}^{B,C,A,k}]$. Case $L_{ij}^{B,C,A,k} < L_{ij}^{C,F,k}$: analogously

The maximizer of (3.7) is denoted $\bar{L}_{ij}^{B,C,k}$, which for all *i*s and *j*s creates a matrix of exposures $\bar{L}^{B,C,k}$.

- (13) [round 3, correction of excessive supply/demand, step k]
 - (a) [for lending] if $\sum_{i} \bar{L}_{ij}^{B,k} > a_j^{B,k}$ then

$$\bar{L}_{ij}^{\mathrm{B},k} := \bar{L}_{ij}^{\mathrm{B},k} \frac{a_j^{\mathrm{B},k}}{\sum_i \bar{L}_{ij}^{\mathrm{B},k}}$$

if
$$\sum_i \bar{L}_{ij}^{\mathrm{B,C},k} > a_j^{\mathrm{C},k}$$
 then

$$\bar{L}_{ij}^{\mathrm{B,C}} := \bar{L}_{ij}^{\mathrm{B,C}} \frac{a_j^{\mathrm{C},k}}{\sum_i \bar{L}_{ij}^{\mathrm{B,C}}}$$

(b) [for borrowing] if $\sum_{j} \bar{L}_{ij}^{B} > l_{i}^{B,k}$ then

$$\bar{L}_{ij}^{\mathrm{B}} := \bar{L}_{ij}^{\mathrm{B}} \frac{l_i^{\mathrm{B},k}}{\sum_i \bar{L}_{ij}^{\mathrm{B}}}$$

if
$$\sum_{j} \bar{L}_{ij}^{\mathrm{B,C}} > l_i^{\mathrm{C},k}$$
 then

$$\bar{L}_{ij}^{\mathrm{B,C}} := \bar{L}_{ij}^{\mathrm{B,C}} \frac{l_i^{\mathrm{C},k}}{\sum_i \bar{L}_{ij}^{\mathrm{B,C}}}$$

www.risk.net/journal

- 62 G. Hałaj et al
 - (14) [update of the network] The networks of interbank $(L^{B,k})$ and nonbank corporate lending $(L^{B,C,k})$ are incrementally updated by the outcomes of the three rounds in step *k*:

$$L^{\mathrm{B},k} := L^{\mathrm{B},k} + \bar{L}^{\mathrm{B},k}$$
$$L^{\mathrm{B},\mathrm{C},k} := L^{\mathrm{B},\mathrm{C},k} + \bar{L}^{\mathrm{B},\mathrm{C},k}$$

The unallocated assets, liabilities and capital are also updated:

$$a_{i}^{B,k+1} = a_{i}^{B,k} - \sum_{j=1}^{N} \bar{L}_{ji}^{B,k}$$

$$a_{i}^{C,k+1} = a_{i}^{C,k} - \sum_{j=1}^{M} \bar{L}_{ji}^{B,C,k}$$

$$l_{j}^{B,k+1} = l_{j}^{B,k} - \sum_{i=1}^{N} \bar{L}_{ji}^{B,k}$$

$$l_{j}^{C,k+1} = l_{j}^{C,k} - \sum_{i=1}^{N} \bar{L}_{ji}^{B,C,k}$$

$$e_{i}^{B,k+1} = e_{i}^{B,k} - \sum_{j=1}^{N} \omega_{j}^{B} \bar{L}_{ji}^{B,k} - \sum_{j=1}^{M} \omega_{j}^{C} \bar{L}_{ji}^{B,C,k}$$

(15) Repeat (10)–(13) until

$$\|L^{\mathbf{B},k+1} - L^{\mathbf{B},k}\| < \epsilon \sum_{i} a_{i} \qquad \text{and} \qquad \|L^{\mathbf{B},\mathbf{C},k+1} - L^{\mathbf{B},\mathbf{C},k}\| < \epsilon \sum_{i} a_{N+i}$$

3.3 Cascade contagion

To assess contagion risk stemming from the interconnectedness on the lending market and a structure of external shocks affecting agents' ability to pay back their debts, we employ a cascade procedure. Let *S* be a vector of real numbers representing a shock structure to credit quality of exposures to individual banks and nonbanks. It is assumed to affect the PDs and is translated to losses reducing banks' capital buffers. Let λ be a vector of loss-given-default (LGD) parameters where its first *N* entries correspond to the interbank exposures and the following *M* to the exposures toward the nonbank corporate sector. Let \mathcal{D}^0 be a set of defaulted banks; that is, those banks for which the capital ratio falls below a certain threshold τ . The initial shock to banks' lending portfolios is calculated as:

$$\Delta_{i}^{e,0} = -\sum_{j=1}^{N} S_{j} \lambda_{j} L_{ji}^{B} - \sum_{j=1}^{M} S_{N+j} \lambda_{N+j} L_{ji}^{B,C}$$

Journal of Network Theory in Finance

www.risk.net/journal

and capital is reduced respectively:

$$e_i^0 = e_i + \Delta_i^{e,0}.$$
 (3.8)

The new capital ratio for the risk-weighted assets R_i is calculated as

$$\mathrm{CR}_i^0 = e_i^0 / (R_i + \Delta_i^{e,0}).$$

For the banks where the capital ratio fall below τ we assume that they default on their interbank obligations, implying that the set of defaulted banks is updated as follows

$$\mathcal{D}^1 = \mathcal{D}^0 \cup \{i \in \bar{N} \mid CR^0_i < \tau\}.$$

Consequently, the cascade is initiated in a sequential way.

Step 1. Let us suppose that a set \mathcal{D}^k , capital position vector e^k and risk-weighted assets vector R^k are known in a certain round k of the cascade.

Step 2. Let a set of new defaults be defined as $\mathcal{D}^{\text{new},k+1} = \{i \in \overline{N} \mid CR_i^k < \tau\}/\mathcal{D}^k$.

- If $\mathcal{D}^{\text{new},k+1} = \emptyset$ then cascade stops.
- If $\mathcal{D}^{\text{new},k+1} \neq \emptyset$ then $\mathcal{D}^{k+1} = \mathcal{D}^k \cup \mathcal{D}^{\text{new},k+1}$ and for all $i \in \overline{N}$

$$\Delta_i^{e,k+1} = -\sum_{j \in \mathcal{D}^{\mathrm{new},k+1}} \lambda_j L_{ji}^{\mathrm{B}}$$

and

$$e_i^{k+1} = e_i^{k+1} + \Delta_i^{e,k+1}, \qquad CR_i^0 = e_i^{k+1} / \left(R_i + \sum_{m=1}^{k+1} \Delta_i^{e,m}\right)$$

and the cascade returns to the beginning of Step 2 for the next round of default calculations (k: = k + 1).

Ultimately, the contagion effects are measured by differences between the terminal capital ratio CR_i^{∞} (after the cascade is unwound) and the starting capital ratio CR_i^0 , ie,

$$\Delta \mathrm{CR}_i = \mathrm{CR}_i^0 - \mathrm{CR}_i^\infty.$$

4 SIMULATIONS

In terms of applied uses the model can be employed for various macroprudential policy analysis purposes to study how contagion can spread within the formed networks and how its effects can be mitigated (by policy actions). Specifically, the generated networks allow for conducting sensitivity analysis of the network structure and the

www.risk.net/journal



FIGURE 4 An example of the estimated network of nonfinancial corporate lending.

contagion risk embedded in this structure to changes in corporate credit risk or in banking sector conditions, including bank regulatory parameters and macroprudential policy instruments (such as capital buffers and large exposure limits).

Different configurations of the network of lending exposures can emerge from the interactions of banks and nonbank corporate firms described by our algorithm.

Source: own calculations, graphs generated using *NetworkX* in Python. The shades of colors of the edges are proportional to the log-exposures. The inner circles represent the banks; the outer circles represent the nonfinancial corporations. The strings attached to the nodes are the short names of the nonfinancial companies; the three-digit codes on the inner circle nodes are the three digits of the EBA bank codes. For clarity, we only keep arrows indicating loans from banks to nonbank agents removing bank-to-bank lending exposures; otherwise numerous bank-to-bank linkages would make the structure of the network invisible.
Randomness of network structures is driven by the probability map and reflects uncertainty about links to be established for a particular pair of firms in the sample of EU banks and borrowers in the EU corporate sector. One particular realization of the distribution of networks generated by the algorithm specified in Section 3.2 is presented in Figure 4 on the facing page. Banks and firms in one country are represented as points on two concentric circles: the inner corresponding to banks and the outer to the nonbank corporate sector. Notably, substantially more links are formed between agents from the same country. However, many larger banks also extend loans to the corporate sector of other countries which can lead to spillovers of shocks originated from corporate subsectors in a country to other counties' banking systems.

To illustrate how the model can be used for policy analysis, we conduct a number of counterfactual simulations of how shocks to different agents may spread to other parts of the system of banks and nonbank firms. For the time being, the simulations mainly focus on how shocks initiated in different parts of the corporate sector propagate to the banking sector via the firms' lending relationships with the banks and are amplified by interbank linkages to a potential cascade of defaults on the interbank market and back again to the real sector (see Section 3.3).

Concretely, we introduce shocks to the creditworthiness of different groups of nonfinancial corporations by increasing their probabilities of default (PDs). An increase in the PD of a specific firm will affect the expected loss (ie, defined as PD times loss-given-default times exposure size) of the credit portfolio of banks holding loans to the firm.

In the first set of counterfactual simulations, we shock the creditworthiness (ie, the PD) of different industry sectors simultaneously across all countries and then use the model to derive the impact of contagion losses of the banks in the network (measured in terms of capital loss of the banks).

Results of this simulation are reported in Figure 5 on the next page. It is observed that there are material differences across sectors in terms of the contagion effects they may inflict on the banking sector (the darker red the columns representing banks in the spectral chart indicate higher contagion-induced losses to bank capital). For example, whereas a shock to the manufacturing sector appears to have a widespread contagion effect on banks throughout the EU, similar shocks to the construction and real estate sectors mainly have material negative implications for banks in Spain, Denmark, Greece and Ireland; countries where banks in recent years (especially pre-crisis) have built up notable exposures to the property developers and construction firms. Shocks to the energy sector are in turn found to produce considerable contagion effects on banks in the Southern European countries, such as Cyprus, Spain, Greece, Italy and Portugal as well as France.

66 G. Hałaj et al



FIGURE 5 Contagion impact on banks: shock to industry sector uniform across all countries.

Values of ΔCR_i for each bank presented; A = agricultural, B = mining, C = manufacturing, D = energy, E = water, sewerage, waste, F = construction, G = trade, H = transportation, I = accommodation/food, J = communication, K = financial, L = real estate. Source: own calculations.

Alternatively, the model can be used to carry out country-specific analysis. For example, what would be the impact of one or more sectors in one country being hit by an adverse shock. This could be analyzed both in terms of the implications for the domestic banking sector and on the cross-border contagion effects to banks in other countries with links to either the industry or to the domestic banks located in the country where the shocks occur.

A first illustration of such simulations is presented in Figure 6 on the facing page which shows the contagion effects on the banking sector when shocking consecutively the PDs of the Spanish industry sectors by ten percentage points. In this particular case, while the domestic banking sector in Spain is affected (especially for shocks affecting the manufacturing, energy and communication sectors) only very limited cross-border contagion is observed. Whereas this may reflect largely nonmaterial cross-border linkages between the Spanish corporate sector and the non-Spanish European banks, the negligible cross-border effects may also be due to the size of the shocks imposed.

To illustrate the latter point, we conducted another set of simulations in which the credit worthiness of the assumedly internationally exposed German manufacturing



FIGURE 6 Contagion impact on banks: ten percentage point shock to PDs of industry sectors in one country, Spain.

Values of ΔCR_i for each bank presented; A = agricultural, B = mining, C = manufacturing, D = energy, E = water, sewerage, waste, F = construction, G = trade, H = transportation, I = accommodation/food, J = Communication, K = financial, L = real estate. Source: own calculations.

sector was shocked to varying degrees. Figure 7 on the next page thus shows the contagion impact on European banks for different sizes of shocks to the PDs of German manufacturing firms. Similar to the Spanish case, for small shocks to PDs the deteriorating credit quality of German manufacturers is found to entail a relatively limited cross-border impact in terms of contagion effects on the banks. However, when gradually increasing the PD shocks it is observed that while the contagion effects within the German sector become more pronounced, also a number of non-German banks start being affected by the shocks to German manufacturing with increasing intensity as the PD shocks become larger. In this example, it is observed that some non-German banks (eg, Spanish, French and Swedish) are affected already for small PD shocks to the German manufacturing sector, whereas broader contagion effects on the European (non-German) banking sector are generated for PD shocks above 60–65%.

The impact of the shocks stemming from the corporate sector may not be limited only to the interbank market but the financial problems of banks may be further transmitted back to the nonbank corporate sector. Such a feedback effect can be related

www.risk.net/journal

68 G. Hałaj et al



FIGURE 7 Contagion impact on banks: shock to the PDs of manufacturing sector in one country, Germany.

Values of ΔCR_i for each bank presented; y axis, different PD levels. Source: own calculations, *Matplotlib* library for Python.

to a reduced provision of financing to the real economy in the case of a defaulting bank or a bank experiencing a substantial decline of the capital has to discontinue extending loans to some sectors of the economy. We illustrate the feedback mechanism referring back to the example of a shock originating in the manufacturing sector in Germany. Hence, Figure 8 on the facing page shows some results of an experiment in which we assumed that banks losing 30% or more of their initial (before the shock) capital cease granting loans to the corporate sector. For a given sector in a country, the feedback effect is quantified by the ratio of loans from failing banks to the total funding needs of the firms in the sector. Figure 8 on the facing page shows that significant feedback effects to the nonbank corporate sector, also with spillovers to corporations in other EU countries, are only found for rather high default probabilities (above 80%). Only in Germany itself can some material disruption to nonbank corporate funding be observed for PD shocks of lower magnitudes as well, eg, around 50%.

While in the above simulations the imposed shocks to the network are all assumed to emerge within the nonbank financial sector, our network model could likewise



FIGURE 8 Reduced provision of credit to the economy by banks affected by a shock to the PDs of manufacturing sector in one country, Germany.

A ratio of the change in funding of nonbank firms to the initial funding volumes (aggregated per country, per sector); x axis, [Country code]_[NACE sector code]; y axis, different PD levels. Source: own calculations, *Matplotlib* library for Python.

be employed to study the reverse causality whereby shocks would occur within the banking sector and spread to other banks and potentially impair corporate lending relationships. Whereas the model is based on micro-level relationships, it could thus also bring insights into macro implications of stress emerging in the banking sector and potentially impairing the banks financial intermediation function. In this context, one way of using the model presented in this paper could be to link it with a macro stress testing framework (see, for example, Bank of England 2013; Henry and Kok 2013) which could be used to derive the initial shocks to banks' capital positions subject to specific macro scenarios. These shocks would then be input into our network model to assess contagion effects to within the banking sector and to the wider economy via the effects on corporate lending relationships.

From a policy perspective, the model-based simulations described in this section and exploiting the fact that the banks in our model are subject to various forms of regulatory constraints (that is, capital constraint, large exposure limits, adjustment of sectoral risk weights, CVA capital charges on counterparty risk) could be useful

www.risk.net/journal

70 G. Hałaj et al



FIGURE 9 Contagion spreading under various regimes of the LE limits.

for assessing the effectiveness of various macroprudential policy actions (see also Hałaj and Kok 2014; Hałaj *et al* 2013). For example, by adjusting minimum capital requirements or large exposure limits imposed on banks in one country or in the whole European Union the model could be used to gauge the effectiveness of these macroprudential policy adjustments by comparing the network-based contagion effects with the contagion in the setting without these adjustments.

We illustrate the application of our framework in the context of large exposure (LE) limit regulations and sector-specific risk weights. Limits are usually set relative to the capital buffers of banks. The commonly used limit in banking regulation, embedded into Basel III regulation framework and its CRDIV legal implementation, assumes that banks are not allowed to keep exposures to one counterparty exceeding 25% of their regulatory capital. Such limits are supposed to increase resilience of banks to defaults of single counterparties but may not be effective in case of multiple defaults within one sector of the economy. The simulations designed to test the sensitivity of contagion losses to various levels of LE limits are presented in Figure 9. In general, as

Values of ΔCR_i for each bank presented; y axis, LE limits; The shock scenario assumes default of 60% of volumes of loans to manufacturing corporate sector in Germany (LGD= 50%). Source: own calculations, *Matplotlib* used for chart generation.



FIGURE 10 Secondary contagion defaults under different regimes of large exposure limits.

Each red (small banks) or blue (large bank) bar indicates a secondary default, ie, banks belonging to the set $s := D^{\infty}/D^1$; *y* axis, LE limits. Source: own calculations, *Matplotlib* used for chart generation.

illustrated by the simulation, the higher the limit is set (in other words, the less stringent regulation is imposed) the higher potential contagion impact of losses originated to the corporate portfolios on banks' capital can be expected. Interestingly, Figure 10 shows that there is a qualitative transition in how the contagion defaults for LE limits above 14% look like. The figure depicts those banks that experience second-round defaults, understood as a default resulting from other banks' defaults and not directly from losses incurred in the corporate loan portfolios that were shocked in the first instance. Formally, the set of banks affected by the secondary defaults (or second-round effects) can be defined in terms of mathematical objects introduced in Section 3.3, ie, as $\delta := \mathcal{D}^{\infty}/\mathcal{D}^1$. For rather restrictive values of LE below 14% no single bank defaults in the simulation as a consequence of any other bank's default. For LE = 15% the number of defaulting banks increases for higher LE limits, particularly visibly for LE limits above 20%. Notably, only smaller banks in the sample are subject to a second round effect except for the case of LE = 25% that creates conditions for a default of one large bank in Germany. Nevertheless, no cross-border default can be observed as an outcome of this particular simulation.

www.risk.net/journal

72 G. Hałaj et al



FIGURE 11 Effect of the increased sectoral risk weights on the magnitude of contagion triggered by banks affected by a shock to the PDs of manufacturing sector in one country, Germany.

Values of ΔCR_i for each bank presented; y axis, different PD levels for (a) risk weights under base case and (b) risk weights increased by a factor of 1.5 for all exposures to the manufacturing sector in Germany. Source: own calculations, *Matplotlib* for Python.

The policy of setting sector-specific risk weights aims to make banks more resilient to risks related to exposures to a certain sector by requiring banks to hold more capital against these exposures. Simulations performed in our framework confirm that this macroprudential policy instrument can be effective in limiting the potential size of the contagion defaults of banks. For illustrative purposes, we assume that the policy maker plans to increase weights for exposures toward the manufacturing sector in Germany. In the next step, it is assumed that an exogenous shock affects the credit quality of banks' loan portfolios exposed to that corporate sector implying credit losses and consequently triggering cascades of interbank defaults. The impact on capital ratios of banks is then compared under two regimes: (a) base case of the estimated risk weight in our model and (b) risk weights to the exposures against the manufacturing sector in Germany increase 1.5 times (eg, from 60% to 90%). Figure 11 shows for different magnitudes of the initial shock to PDs that higher weights result in lower losses. This finding is especially pronounced for very strong shocks of PD > 80%.

Journal of Network Theory in Finance

www.risk.net/journal

5 CONCLUSIONS

In this paper, we extend the interbank network analysis that is standard in the literature to also encompass firm-specific networks. This allows us to study risks stemming from interconnectedness not only within the financial system but also those arising due to interlinkages between the financial and real sectors of the economy. In another extension to the standard literature, we furthermore apply agent-based network formation methods. This inter alia allows us to examine how changes in key prudential parameters, such as minimum capital requirements and large exposure limits, can influence the contagion risks embedded in the network structures as a function of the (constrained) optimizing behavior of the agents in the network.

The simulations presented in the paper clearly demonstrate that also adding network structures capturing bank–firm lending relationships is a relevant layer of interconnectedness that can amplify the contagion risks that in the literature have typically been confined to within-financial sector networks.

Furthermore, the model can also be a useful analytical tool to study second-round effects in relation to bank stress tests whereby the "first round" stress imposed on the banking sector within the framework can propagate beyond the banks via their lending relationships with the corporate sector.

Future extensions of the model should focus on bringing into the network further economic sectors, such as households, governments and other financial intermediaries. Especially, embedding a network of households will require a different type of data set than the one we have employed in this study and we will leave this for future research.

DECLARATION OF INTEREST

This research document should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

ACKNOWLEDGEMENTS

We are grateful for very valuable comments received from participants of the "Systemic Risk: Models and Mechanisms" workshop at Newton Institute (Cambridge, 2014), the "Financial Risk and Network Theory" seminar at Judge Business School (Cambridge, 2014), the CFE-ERCIM 2014 Annual conference (Pisa, 2014) and the CEMLA Seminar on Network Analysis and Financial Stability Issues (Mexico City, 2014).

www.risk.net/journal

74 G. Hałaj et al

REFERENCES

- Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2013). Systemic risk and stability in financial networks. Working Paper 18727, National Bureau of Economic Research, January. URL: www.nber.org/papers/w18727.
- Adam, A. (2008). Handbook of Asset and Liability Management: From Models to Optimal Return Strategies. The Wiley Finance Series. Wiley
- Aldasaro, I., Delli Gatti, D., and Faia, E. (2014). Bank networks: contagion, systemic risks and prudential policy. Mimeo, European Central Bank.
- Allen, F., and Babus, A. (2009). Networks in finance. In *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*, Kleindorfer, P., and Wind, J. (eds). Wharton School Publishing
- Allen, F., and Gale, D. (2000). Financial contagion. *Journal of Political Economy* **108**(1), 1–33.
- Angelini, P., Di Salvo, R., and Ferri, G. (1998). Availability and cost of credit for small businesses: customer relationships and credit cooperatives. *Journal of Banking and Finance* **6–8**(22), 925–954.
- Aoyama, H., Ikeda, Y. Iyetomi, H. Fujiwara, Y., and Souma, W. (2009). Structure and temporal change of the credit network between banks and large firms in Japan. *Open-Assessment EJournal* **3**(7).
- Babus, A. (2011). Endogenous intermediation in the over the counter markets. Working Paper, Imperial College London.
- Berger, A. N., and Udell, G. F. (1995). Relationship lending and lines of credit in small firm finance. *Journal of Business* **68**, 351–381.
- Bank of England (2013). A framework for stress testing the UK banking system. Discussion Paper, October.
- Bank for International Settlements (2013). Macroeconomic impact assessment of the OTC derivatives reforms. Technical Report, Bank for International Settlements, August. URL: www.bis.org/press/p130826.htm
- Blasques, F., Koopman, S. J., Lucas, A., and Schaumburg, J. (2014). Spillover dynamics for systemic risk measurement using spatial financial time series models. Working Paper 16, SYRTO.
- Bluhm, M., Faia, E., and Krahnen, P. (2013). Endogenous banks' networks, cascades and systemic risk. Mimeo, University of Frankfurt, March. URL: www.wiwi.uni-frankfurt.de/ profs/faia/.
- Boot, A. W. A. (2000). Relationship banking: what do we know? *Journal of Financial Intermediation* **9**(1), 7–25.
- Boot, A. W. A., and Thakor, A. V. (2000). Can relationship banking survive competition? *Journal of Finance* 55(5), 679–713.
- Boss, M., Elsinger, H. Thurner, S., and Summer, M. (2004). Network topology of the interbank market. *Quantitative Finance* **4**,1–8.
- Canedo, J. M. D., and Martínez-Jaramillo, S. (2009). A network model of systemic risk: stress testing the banking system. *Intelligent Systems in Accounting, Finance and Management* **16**(1–2), 87–110.

- Castiglionesi, F., and Lavarro, N. (2011). Fragile financial networks. Mimeo, Tilburg University.
- Castiglionesi, F., and Wagner, W. (2013). On the efficiency of bilateral interbank insurance. *Journal of Financial Intermediation*, forthcoming.
- Chen, J., and Song, K. (2013). Two-sided matching in the loan market. *International Journal* of *Industrial Organization* **31**, 145–152.
- Cohen-Cole, E., Patacchini, E., and Zenou, Y. (2011). Systemic risk and network formation in the interbank market. Discussion Paper 8332, CEPR, April.
- Cesar de Castro Miranda, R., and Miranda Tabak, B. (2013). Contagion risk within firm-bank bivariate networks. Working Paper 322, Banco Central de Brazil.
- Duffie, D., and Sun, Y. (2012). The exact law of large numbers for independent random matching. *Journal of Economic Theory* **147**, 1105–1139.
- European Central Bank (2008). The role of banks in the monetary policy transmission mechanism. Monthly bulletin article, European Central Bank, August.
- European Central Bank (2012). Heterogeneity in euro area financial conditions and policy implications. Monthly bulletin article, European Central Bank, August.
- Eisenschmidt, J., and Tapking, J. (2009). Liquidity risk premia in unsecured interbank money markets. Working Paper 1025, European Central Bank,.
- Furfine, C. (2003). Interbank exposures: quantifying the risk of contagion. *Journal of Money, Credit and Banking* **35**(1), 111–638, .
- Gallegati, M., Greenwald, B. G. De Masi, Fujiwara, Y., and Stiglitz, J. E. (2010). An analysis of the Japanese credit networks. *Evolutionary and Institutional Economics Review* **7**(2), 209–232.
- Georg, C.-P. (2013). The effect of the interbank network structure on contagion and common shocks. *Journal of Banking and Finance* **37**(7), 2216–2228.
- Grasselli, M. (2013). An agent-based computational model for bank formation and interbank networks. In *Handbook on Systemic Risk*, Fouque, J.-P., and Langsam, J. (eds), Cambridge University Press.
- Hałaj, G. (2012). Systemic bank valuation and interbank contagion. In Advances in Network Analysis and Applications, Kranakis, E. (ed). Proceedings of MITACS Workshops. Springer Series on Mathematics in Industry.
- Hałaj, G., and Kok, C. (2013). Assessing interbank contagion using simulated networks. *Computational Management Science* **10**(2–3), 157–186.
- Hałaj, G., and Kok, C. (2014). Modeling emergence of the interbank networks. *Quantitative Finance*. 10.1080/14697688.2014.968357.
- Hałaj, G., Kok, C., and Montagna, M. (2013). Gauging the effectiveness of cross-sectional macroprudential tools through the lens of interbank networks. European Central Bank Financial Stability Review, Special Feature C, November.
- Henry, J., and Kok, C. (2013). A macro stress testing framework for assessing systemic risks in the banking sector. Occasional Paper 152, European Central Bank.
- Iori, G., De Masi, G.Precup, O. V. Gabbi, G., and Caldarelli, G. (2008). A network analysis of the Italian overnight money market. *Journal of Economic Dynamics and Control* 32(1), 259–278.
- Jackson, M. O., and Watts, A. (2010). Social games: matching and the play of finitely repeated games. *Games and Economic Behavior* **70**(1), 170–191.

www.risk.net/journal

- 76 G. Hałaj et al
 - Lopez-Castañon, C., Martínez-Jaramillo, S., and Lopez-Gallo. F. (2012). Systemic risk, stress testing, and financial contagion: their interaction and measurement. In *Simulation in Computational Finance and Economics: Tools and Emerging Applications*, Biliana Alexandrova-Kabadjova, B. and Martinez-Jaramillo, S. (eds), pp. 181–210. IGI Global.
 Manage M. (2014). Intermediation in paturation. Martinezer Ma
 - Manea, M. (2014). Intermediation in networks. Working Paper, MIT.
 - Markose, S. (2012). Systemic risk from global financial derivatives: a network analysis of contagion and its mitigation with super-spreader tax. Working Paper 12/282, IMF.
 - De Masi, G., and Gallegati, M. (2012). Bank–firms topology in italy. *Empirical Economics* **43**,851–866.
 - Montagna, M., and Kok, C. (2013). Multi-layered interbank model for assessing systemic risk. Kiel Working Papers 1873, Kiel Institute for the World Economy.
 - Nier, E., Yang, J., Yorulmazer, T., and Alentorn, A. (2007). Network models and financial stability. *Journal of Economic Dynamics and Control* **31**(6), 2033–2060.
 - Petersen, M. A., and Rajan, R. G. (1994). The benefits of lending relationships: evidence from small business data. *Journal of Finance* **49**, 3–37.
 - Soramaki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A* 379, 317–333.
 - t'Veld, D., van der Leij, M., and Hommes, C. (2014). The formation of a core periphery structure in heterogeneous financial networks. Discussion Papers 098-II, Tinbergen Institute.
 - Upper, C., and Worms, A. (2004). Estimating bilateral exposures in the German interbank market: is there a danger of contagion? *European Economic Review* **48**(4), 827–849.
 - van Lelyveld, I., and Liedorp, F. (2006). Interbank contagion in the Dutch banking sector: a sensitivity analysis. *International Journal of Central Banking* **2**(2), 99–134.



Research Paper

Risk diversification: a study of persistence with a filtered correlation-network approach

Nicoló Musmeci¹, Tomaso Aste^{2,3} and T. Di Matteo¹

¹Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK; emails: nicolo.musmeci@kcl.ac.uk, tiziana.di_matteo@kcl.ac.uk
²Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK; email: t.aste@ucl.ac.uk
³Systemic Risk Centre, London School of Economics and Political Sciences, Houghton Street, London WC2A 2AE, UK

(Received October 20, 2014; revised December 17, 2014; accepted January 9, 2015)

ABSTRACT

In this paper, we study the evolution over time of the correlation structure of equity returns by means of a filtered-network approach and use this to investigate persistency and recurrences and their implications for risk-diversification strategies. We build dynamically planar maximally filtered graphs from the correlation structure over a rolling window and study the persistence of the associated directed bubble hierarchical tree (DBHT) clustering structure. We observe that the DBHT clustering structure is quite stable during the early 2000s, becoming gradually less persistent before the unfolding of the 2007–8 crisis. The correlation structure eventually recovers persistence in the aftermath of the crisis, setting up a new phase, distinct from the precrisis structure, in which the market structure is less related to industrial sector activity. We observe that the correlation structure is again losing persistence, which indicates the building up of another, different phase. Such dynamical changes in persistency and its occurrence at the unfolding of financial crises raise concerns about the effectiveness of correlation-based portfolio management tools for risk diversification.

Corresponding author: T. Di Matteo

Print ISSN 2055-7795 | Online ISSN 2055-7809 Copyright © 2015 Incisive Risk Information (IP) Limited

Keywords: econophysics; correlation-based filtered networks; information filtering networks; systemic risk; risk diversification; community tracking.

1 INTRODUCTION

One way to reduce financial risk is by diversifying investments. This involves taking positions in assets that are historically anticorrelated or uncorrelated and in this way reducing the probability that all assets lose value at the same time. This is, for instance, the basis of the capital asset pricing model (Fama and French 2004). However, the applicability of these approaches relies on the implicit assumption that the relevant features of the correlation structure observed in the past have persistent significance into the future. This is not always the case.

In order to characterize the correlation structure and quantify its persistence, we use a network-filtering approach in which the correlation matrix is mapped into a sparse graph that retains only the relevant elements. To this purpose, we use the correlationfiltered networks known as planar maximally filtered graphs (PMFGs) (Tumminello *et al* 2005) and their associated clustering structure, the directed bubble hierarchical tree (DBHT) (Song *et al* 2012). A PMFG is a maximal planar graph that retains only the largest correlations. The DBHT is a hierarchical clustering that is constructed by making use of the separating properties of 3-cliques in planar graphs (Song *et al* 2011).

Since the seminal work of Mantegna (1999), network analysis on asset correlation has provided interesting insights into risk management and portfolio optimization. It has been observed that the structure of such networks not only is significantly related to the industrial sectors' classifications, but also conveys important independent information (Mantegna 1999; Musmeci *et al* 2014). It was shown in Borghesi *et al* (2007) that this network structure can be very robust against changes in the time horizon at which the asset returns are sampled (when the market mode dynamics is removed from the original correlations). This has been interpreted as an indication that "correlations on short time scales might be used as a proxy for correlations on longer time horizons" (Borghesi *et al* 2007). This, however, requires some degree of stationarity in the correlation structure.

Network-filtering procedures have been found to sensibly improve the performance of portfolio optimization methods. For instance, it was shown in Tola *et al* (2008) that Markowitz optimization gives better results on network-filtered correlation matrixes than on unfiltered ones. In Pozzi *et al* (2013), it was reported that the peripheral position of nodes in PMFGs can be a criterion for selecting a well-diversified portfolio. This finding is consistent with that for the minimum spanning tree (MST) in Onnela

et al (2003b): the stocks selected using the Markowitz method tend to be the "leaves" of the MST.

Network-filtered correlations carry both local and global information in their structures and the analysis of their temporal evolution may allow us to better understand financial market evolution. For instance, Di Matteo et al (2010) observed that stocks belonging to the same industrial sector tend to have similar values of centrality in the network topology, and that this differentiation is persistent over time. In particular, they observed that finance, basic materials and capital goods industrial sectors (Forbes classification) tend to be located mostly in the central region of the network, whereas energy, utilities and health care are located more in the peripheral region. The preeminent role of the financial sector is even stronger when correlation networks based on partial correlations are analyzed (Kenett et al 2010). Despite this overall robustness, a certain degree of nonstationarity has also been observed. For instance, the financial sector appears to have lost centrality over the first decade of the 2000s (Aste et al 2010). Buccheri et al (2013) found both a slow and a fast dynamics in correlation networks topology: while the slow dynamics shows persistence over periods of at least five years, the time scale of the fast dynamics is a few months and linked to special exogenous and endogenous events, such as financial crises. For instance, Onnela et al (2003a) showed that sharp structural changes occurred in the graph topology during Black Monday 1987. Similar phenomena have been observed for correlations on foreign exchange (FX) data (Jang et al 2010). McDonald et al (2008) demonstrated that structural changes on FX correlation data display different features depending on the type of event affecting the market. News that concerns economic matters can trigger a prompt destabilizing reaction, whereas there are periods of "collective discovery" in which dynamics appears to synchronize (McDonald et al 2008).

In this paper, we investigate the nonstationarity of correlation, quantifying how much, and in what way, the correlation structure changes over time. This is a particularly relevant topic because most portfolio optimization tools rely on some stationarity, or at least persistence, in the joint distribution of asset returns. It is generally accepted in the literature that financial correlations are nonstationary. For instance, in Livan *et al* (2012) it was shown, by means of a local Kolmogorov–Smirnov test on correlation pairs, how nonstationarity can sensitively affect the effectiveness of portfolio optimization tools. In this paper, we discuss the degree of nonstationarity in the correlations at a nonlocal level by using PMFG networks and the associated DBHT clustering and looking at changes in the hierarchical and clustering structures. In this context, persistence translates into a measure of similarity among communities in a network, for which network-theoretic tools should be used. The PMFG–DBHT method has recently been applied to the study of financial data (Musmeci *et al* 2014), showing that it is a powerful clustering tool that can outperform other traditional clustering methods, such as linkage and *k*-medoids, in retrieving economic information.

Moreover, the dynamical analyses have shown that the clustering structure reveals peculiar patterns over the financial crisis, for instance, the increasingly dominant role of the market mode over the period 1997–2012. This implies an increase in nondiversifiable risk in the market. In this paper, we take these analyses a step further by looking at the dynamics of this clustering and its persistence.

The rest of the paper is organized as follows. In Section 2, we summarize the main theoretical concepts underlying the correlation network tools. In Section 3, we describe the analyses we have performed and discuss the results. In Section 4, we draw our conclusions and discuss future perspectives.

2 CORRELATION-BASED NETWORKS: AN OVERVIEW

Over the last fifteen years, correlation-based networks have been used extensively in the econophysics literature as tools to filter and analyze financial market data (Aste *et al* 2005; Bartolozzi *et al* 2007; Di Matteo and Aste 2002; Di Matteo *et al* 2004, 2005; Mantegna 1999; Onnela *et al* 2003c; Tumminello *et al* 2005).

The seminal work of Mantegna (1999) exploited for the first time a tool from network theory, the MST (see, for example, West 1996), to analyze and filter from noise the correlation structure of a set of financial assets. Mantegna's idea was to look at a correlation matrix as the adjacency matrix of a network and generate an MST on this network in order to retain the most significant links/entries. Moreover, after mapping the correlation into a suitable metric distance, the MST algorithm provides a hierarchical classification of the stocks.

In the following years, other correlation-based networks were studied in the literature. Onnela *et al* (2003c) introduced the dynamic asset graph. Unlike the MST, which filters the correlation matrix according to a topological constraint (the tree-like structure of the MST), the dynamic asset graph retains all the links, such that the associated correlation (distance) is above (below) a given threshold. In this way, it is less affected by the insignificant, low correlations that are often kept by the MST. As a result, the dynamic asset graph is more robust against time (Onnela *et al* 2003c). On the other hand, the MST, by retaining both high and low correlations, is better equipped to uncover global, multiscale structures of interaction. Indeed, in financial and complex systems in general, several length scales coexist, and thresholding at a given value artificially introduces a characteristic size that might hide effects occurring at other scales.

The tree structure exploited in the MST tool is not the only topological constraint that can be used to filter information. In particular, if we replace the request of absence of loops with the planarity condition, we obtain the PMFG (Aste *et al* 2005). The PMFG can be seen as a generalization of the MST that is able to retain a greater amount of information (Aste 2012; Tumminello *et al* 2005), as it has a less strict

topology constraint that allows it to keep a larger number of links. It can be shown that the hierarchical properties of the MST are preserved in the PMFG.

We can take this concept a step further and generalize the PMFG to a broader class of networks by means of the concept of "genus" (Aste et al 2005). The genus, g, of a surface is the largest number of nonintersecting simple closed cuts that can be made on the surface without disconnecting a portion (equal to the number of handles in the surface). Requiring a network to be planar, as for the PMFG, is equivalent to requiring that the network be embedded on a surface with g = 0 (ie, no handles, a topological sphere). The natural generalization of the PMFG is therefore a network embedded on surfaces with genus greater than zero. The greater the genus, the more handles are in the surface and the more links we can retain from the original correlation matrix. More links retained means more information and network complexity, but it also means more noise. When $g = \left[(N-3)(N-4)/12 \right]$ (where N is the number of nodes and $\lceil x \rceil$ is the ceiling function that returns the smallest integer greater than or equal to x), the original, fully connected, complete graph associated with the correlation matrix can be recovered. The concept of embedding on surfaces therefore provides a quantitative way of tuning the degree of information filtering by means of a single parameter, g, linking correlation-based networks to algebraic geometry (Aste et al 2012).

Correlation-filtered networks are associated with clustering methods. Indeed, the MST is strictly related (Tumminello *et al* 2010) to a hierarchical clustering algorithm, namely the single linkage (SL) (Anderberg 1973). MST can indeed be seen as a network representation of the hierarchy generated by the SL. Recently, it has been shown that a hierarchical clustering can be derived from the PMFG as well (Aste 2014; Song *et al* 2012). This new method is the DBHT. However, the approach is different from the agglomerative one adopted in the linkage methods: the idea of the DBHT is to use the hierarchy hidden in the topology of a PMFG, due to its being made of 3-cliques (Song *et al* 2011, 2012). The DBHT hierarchical clustering was applied to synthetic and biological data in Song *et al* (2012) and financial data in Musmeci *et al* (2014), showing that it can outperform several other clustering methods, including *k*-means++ (Arthur and Vassilvitskii 2007), *k*-medoids (Kaufman and Rousseeuw 1987), linkage, spectral clustering via normalized cuts on *k*-nearest neighbor graphs (*k*NN-spectral) (Shi and Malik 2000), the self-organizing map (SOM) (Kohonen *et al* 2001) and the Qcut (Ruan *et al* 2010).

Since the DBHT exploits the topology of the correlation network, it can be viewed as an example of community-detection algorithms in graphs (Fortunato 2010). The implicit assumption underlying these algorithms is that a community is somehow related to the density of edges inside and outside the community itself, unlike strict data-clustering methods (such as the aforementioned linkage algorithms), which only

use the information contained in the similarity/distance matrix. Several different community detection algorithms in graphs have been suggested in the literature; many of them search for the community partition that maximizes modularity, a function that compares the density of links in each community with the one expected in a (null) random graph model (Guimerà *et al* 2004; Newman 2004; Newman and Girvan 2004). Other approaches include spectral analysis on the adjacency matrix or related matrixes, eg, Laplacian (Donetti and Muñoz 2004; Mitrović and Tadić 2009), random walks on networks (Hu *et al* 2008; Zhou 2003) and methods based on statistical inference (Reichardt and White 2007; Rosvall and Bergstrom 2007).

However, in this paper we focus on the DBHT, as it is tailored to planar graphs and is therefore the natural tool to use with PMFGs.

3 PERSISTENCE AND TRANSITIONS: DYNAMICAL ANALYSIS OF THE DIRECTED BUBBLE HIERARCHICAL TREE

We studied the dynamical evolution of DBHT clustering on a system of N = 342 US stocks during the time period January 1997–December 2012. We selected a set of n = 100 overlapping time windows, T_k , with k = 1, ..., n (each one of length L = 1000 trading days with a shift of 30 trading days between adjacent time windows), and computed the distance matrix

$$D_{ij}(T_k) = \sqrt{2(1 - \rho_{ij}(T_k))},$$

where ρ_{ij} is the Pearson correlation coefficient

$$\rho_{ij}(T_k) = \frac{\langle c_i(t)c_j(t)\rangle_{T_k}}{\sqrt{[\langle c_i^2(t)\rangle_{T_k} - \langle c_i(t)\rangle_{T_k}^2][\langle c_j^2(t)\rangle_{T_k} - \langle c_j(t)\rangle_{T_k}^2]}},$$
(3.1)

where $\langle \cdot \rangle_{T_k}$ represents the average over the time window T_k , and $c_i(t)$, $c_j(t)$ are the daily log returns of stocks *i* and *j* detrended of the average market return factor. Following Borghesi *et al* (2007), we computed $c_i(t)$ for each stock *i*, assuming the following one-factor model for the stock log return $r_i(t)$:

$$r_i(t) = \alpha_i + \beta_i I(t) + c_i(t), \qquad (3.2)$$

where the common market factor I(t) is the market average return,

$$I(t) = \sum_{\gamma=1}^{N} r_{\gamma}(t).$$

The coefficients α_i , β_i are computed by means of a linear regression and $c_i(t)$ is the residual. In agreement with Borghesi *et al* (2007), we verified that correlations



FIGURE 1 Dynamical evolution of the DBHT clustering.

Each plot refers to 100 moving time windows (T_k) of length 1000 trading days and shift 30 days. (a) Number of DBHT clusters, N_{cl} , with the dashed horizontal line representing the N_{cl} value obtained by taking the entire time window of 4026 trading days (covering years 1997–2012). Overall, we can observe a drop in correspondence with the 2007–2008 financial crisis. (b) Clustering similarity with industrial classification benchmark (ICB) classifications. This graph shows the amount of economic information retrieved by DBHT clustering in terms of similarity between clustering and ICB partitioning, calculated using the adjusted Rand index, \Re_{adj} . Again, a drop at the outbreak of the crisis appears. Over the postcrisis years, there is less economic information than in the precrisis period, and differences among different ICB levels are less evident. (c) Similarity between consecutive clustering, showing the persistence of the DBHT clustering over time, measured as the adjusted Rand index between two adjacent clusterings. The financial crisis is characterized by very low levels of persistence.

on detrended log returns provide a richer and more robust clustering that can carry information not evident in the original correlation matrix (Borghesi *et al* 2007). We also used a weighted version of the Pearson estimator (Pozzi *et al* 2012) in order to mitigate (exponentially) excessive sensitiveness to outliers in remote observations. The DBHT clustering is calculated on each distance matrix $D(T_k)$.

In part (a) of Figure 1, we show the number of DBHT clusters obtained for each time window. The number of clusters ranges from 14 to 26. The dashed line is the value

www.risk.net/journal

(23) corresponding to the clustering obtained using the entire period 1997–2012 as our time window. As we can observe, the lowest values are associated with the period around the 2007–8 financial crisis.

In order to analyze the amount of economic information expressed by the clustering (Coronnello *et al* 2011; Mantegna 1999), we measured the adjusted Rand index, \mathcal{R}_{adj} (Hubert and Arabie 1985), between the DBHT clustering at time window T_k and the community partition generated by the industrial sector classification of stocks. \mathcal{R}_{adi} is an index that measures the similarity between two different partitions on the same set of objects (stocks in this case) and ranges from 0 (no similarity) to 1 (complete identity). We provide a formal definition of this index in Appendix A. \mathcal{R}_{adj} therefore provides a measure of the industrial information contained in the correlation-based clustering. We use the industrial classification benchmark (ICB), which is a categorization that divides the stocks into four hierarchical levels: namely, 114 subsectors, 41 sectors and 19 different supersectors (which, in turn, are gathered in ten different industries). In order to take all of these levels into account, we measured $\mathcal{R}_{adi}(T_k)$ between each of the hierarchical levels and DBHT clustering. In part (b) of Figure 1 on the preceding page, we plot the evolution over time of $\mathcal{R}_{adj}(T_k)$ between the DBHT clusters and ICB industries, supersectors and subsectors (for simplicity, we do not plot sector data that is very close to supersectors' values). We can see how the ICB information shows a remarkable drop during the 2007–8 financial crisis, which partially recovers from 2010 onward. Interestingly, before the crisis, the industry, supersector and subsector lines were distinct (with ICB supersectors showing the highest similarity with the DBHT, followed by industries and subsectors), whereas in the crisis and postcrisis periods they display much closer values. Therefore, from the crisis onward correlation clustering is no longer able to distinguish between different levels of ICB. This might indicate that this industrial classification is becoming a less reliable benchmark to diversify risk. These results are confirmed by other industrial partitions, including the Yahoo classification.

The adjusted Rand index can also be used as a tool for analyzing the persistence of the DBHT clustering by measuring the index between two clusterings at two adjacent time windows (we denote by $\mathcal{R}_{adj}^{T-1,T}(T_k)$ such a quantity). This gives a measure of local persistence: a drop in the index value indicates decreasing similarity between adjacent clusterings, and therefore less persistence. In part (c) of Figure 1 on the preceding page, we plot $\mathcal{R}_{adj}^{T-1,T}(T_k)$ against time. We observe that the clustering persistence changes remarkably over time, dropping in particular with the outbreak of the financial crisis and recovering in 2010. Note that the drop during the crisis starts earlier than the actual outbreak of it (August 2007, the dashed vertical line). This could highlight a possible use of clustering persistence as a tool to forecast systemic risk. Notably, in 2010–12 we again observe a steadily decreasing trend. Interestingly, the pattern of persistence appears to be related to the similarity between clustering and ICB, with periods of higher persistence characterized by higher amounts of economic information.

However, the drawback of $\mathcal{R}_{adj}^{T-1,T}(T_k)$ as a measure of persistence is that, at any one time, it only provides information on the persistence with respect to the previous, adjacent time window. It tells us nothing about the long-term robustness of each clustering. To investigate this aspect, in Section 3.1 we discuss a set of analyses that evaluate the persistence of each clustering at each time, therefore providing a more complete picture.

3.1 A map of structural changes

To investigate the long-term persistence of each clustering, we calculated the adjusted Rand index for each time window between the corresponding clustering and the clustering at any other time. The result is summarized in the (symmetric) similarity matrix *s*:

$$s(T_a, T_b) = \mathcal{R}_{\text{adj}}(X_a, X_b), \tag{3.3}$$

where X_a and X_b are the DBHT clusterings at time windows T_a and T_b respectively. The matrix *s* for our data set is shown in part (a) of Figure 2 on the next page. We observe two main blocks, the first precrisis and the other postcrisis, within which high similarity among clusterings may be found. The two blocks show very low mutual similarity (upper right corner/lower left corner of the matrix). The first block begins losing its compactness in 2007, and the second block quite quickly does the same at the beginning of 2011. Between these two times, the outbreak of the financial crisis displays a series of extremely changeable clusterings that do not show similarity with any other time window.

To better highlight these changes of regime, we plot in parts (b)–(e) of Figure 2 on the next page four time rows from matrix *s*, taken as examples of persistence behavior during the precrisis (part (b) September–October 2003), crisis (part (c) July–August 2007, the outbreak of the crisis, and part (d) November–December 2008, the aftermath of Lehman Brothers' default) and postcrisis periods (part (e) April–May 2010). The vertical dashed lines show the end position of the time window whose clustering is taken as a reference. Each point in the plot is the adjusted Rand index between that clustering and all the other clusterings at each other time window, in both the past and the future. In the precrisis period (b), the similarity displays a quite slow decay both forward and backward in time; the original clustering still has a 60% similarity with the seventeenth time window forward/backward in time. The decreasing trend is, however, evident and becomes steeper during the crisis. Taking time windows during the financial crisis, (c) and (d), the pattern changes drastically: the similarity drops by 70–80% in a few months both backward and forward in time. The two stages of crisis also reveal some differences. While in the early crisis period (c) the similarity with





(a) Similarity matrix *s* showing the temporal evolution of the correlation-based DBHT clustering. Each entry $s(T_a, T_b)$ is the adjusted Rand index between clustering X_a and X_b at time windows T_a and T_b respectively (3.3). Higher values indicate greater similarity. The matrix displays two main blocks of high intrasimilarity: one precrisis and the other postcrisis. The years 2007–2008 fall between these two blocks and display very low similarity with any other time window, revealing an extremely changeable structure. Parts (b)–(e) show the patterns of similarity for four sample time windows (ie, four sample rows of the similarity matrix): (b) September–October 2003, (c) July–August 2007, (d) November–December 2008 and (e) April–May 2010. During the crisis, similarity decays much faster than in the precrisis and postcrisis periods.

Journal of Network Theory in Finance

www.risk.net/journal



FIGURE 3 Persistence analysis based on metacorrelation.

(a) Similarity matrix *z* showing the temporal evolution of correlation matrixes. Each entry $z(T_a, T_b)$ is calculated as a correlation among correlation matrixes at time windows T_a and T_b (3.4). Higher values indicate higher similarity. Parts (b)–(e) show the patterns of similarity for four sample time windows: (b) September–October 2003, (c) July–August 2007, (d) November–December 2008 and (e) April–May 2010. The decay during the crisis years is much less steep than in the corresponding plot in Figure 2 on the facing page.

www.risk.net/journal

precrisis clusterings is higher than with postcrisis ones, in the post-Lehman Brothers period (d) the situation is reversed. Finally, the postcrisis period (e) shows a partially recovered persistence, although not at the same levels as the 2003 pattern.

We may wonder whether the structural changes highlighted by the clustering analyses can be detected directly by studying the original, unfiltered correlation matrixes. To check this, we introduce an alternative measure of similarity among different time windows that does not make any use of clustering: namely, the correlation between the coefficients of two correlation matrixes (metacorrelation). This measure is

$$z(T_a, T_b) = \frac{\langle \rho_{ij}(T_a)\rho_{ij}(T_b)\rangle_{ij}}{\sqrt{[\langle \rho_{ij}^2(T_a)\rangle_{ij} - \langle \rho_{ij}(T_a)\rangle_{ij}^2][\langle \rho_{ij}^2(T_b)\rangle_{ij} - \langle \rho_{ij}(T_b)\rangle_{ij}^2]}}, \quad (3.4)$$

where $\rho_{ii}(T_a)$ is the correlation between stocks *i* and *j* at time window T_a and $\langle \cdot \rangle_{ii}$ is the average over all couples of stocks i, j. Munnix et al (2012) introduced an alternative measure to identify the possible states of a financial market. In Figure 3 on the preceding page, we report the matrix $z(T_a, T_b)$ and four representative time rows, which correspond to the same four time windows chosen in Figure 2 on page 86. We observe that metacorrelation is indeed able to identify the two precrisis and postcrisis time blocks. However, it also shows a smaller, intermediate block during the 2007– 8 crisis with a relatively high intrasimilarity. This is different than what we have observed in the clustering-based matrix s, where the time windows during the crisis were quite dissimilar. Moreover, the precrisis and postcrisis blocks in z display higher intrasimilarity than s, especially over the postcrisis years. All these differences can be appreciated when looking at the four z time rows in parts (b)–(e) of Figure 3. Even if in the crisis time windows (c) and (d) we observe a faster decay of similarity, it is much less steep than the corresponding clustering plot (parts (c) and (d) of Figure 2). Moreover, the postcrisis window in part (e) of Figure 3 recovers completely the high precrisis level of persistence, unlike the clustering case in part (e) of Figure 2.

Therefore, it seems that metacorrelation and clustering analyses depict different dynamics of market correlation structure. In particular, the clustering-based matrix *s* reveals higher nonstationarity during the crisis and postcrisis periods. The instability of correlation during crises has recently been observed by Chetalova *et al* (2014); however, their result relies on a specific choice for the multivariate distribution of returns, whereas our analyses are model independent.

3.2 Clusters composition evolution

So far, we have described the persistence of clusters from a global perspective, looking at the clustering as a whole. Let us here focus on the evolution of each cluster, following how their composition changes over time. It is not straightforward to analyze such an evolution, with the main problem being the changeable nature of dynamical clusters, which makes it difficult to identify the successor for each cluster. Many different approaches can be adopted to address this community tracking problem (Fenn *et al* 2012). Here, we use hypothesis statistical tests based on the hypergeometric distribution (Feller 2008; Tumminello *et al* 2011a) to assess similarity between clusters at different times. In particular, if the number of stocks in common between two clusters is high enough to reject the null hypothesis of the test, we label the two clusters as "similar". Moreover, we take the DBHT clustering calculated over the entire time window (1997–2012) as a benchmark clustering through which we can track the evolution of the dynamical clusters obtained with the moving time windows. Let us here describe this idea in more detail.

Let us call X the clustering obtained on the entire time window and Y_i a cluster belonging to X, with $i = 1, ..., N_{cl}$. For each cluster Y_i , and for each time window T_k (k = 1, ..., n), we have taken the clustering at time T_k , X_{T_k} , and identified the cluster belonging to X_{T_k} that is "similar" to Y_i (if any). We label a cluster as "similar" to Y_i if the number of stocks in common with Y_i is high enough to reject the null hypothesis of the hypergeometric test (Musmeci *et al* 2014; Tumminello *et al* 2011b). This test considers a random overlapping between the two clusters (a detailed description of the test can be found in Appendix B). If more than one cluster turns out to be similar, we take the largest cluster. Eventually, we end up with one cluster for each Y_i for each time window T_k . All of them have a high degree of similarity with Y_i in common. We can therefore follow their evolution in terms of the number of stocks and corresponding ICB industrial sectors. The threshold for the tests was chosen equal to 0.01, together with the conservative Bonferroni correction (Feller 2008).

In part (a) of Figure 4 on the next page, the composition of the DBHT clustering computed over the time window 1997–2012 is shown. For each cluster, the *y*-axis displays its cardinality S (ie, the number of stocks belonging to the cluster), with different shades showing stocks belonging to different ICB industries. For the eleven biggest clusters in X, we plot in parts (b)–(f) of Figure 4 and parts (a)–(f) of Figure 5 on page 91 the number of stocks S for their similar clusters in time, together with their composition in terms of ICB industries. When no similar clusters can be found for a time window, we have just left the correspondent window empty. The clusters analyzed are the numbers 18, 4, 8, 7, 17, 1, 6, 20, 14, 22 and 15. We summarize the main findings below.

• Overall, all the clusters in X have a high persistence over time, showing a corresponding "similar" cluster at almost every time window. This result is remarkable, as the persistence has been assessed in quite a conservative way, ie, using the hypergeometric test with a Bonferroni correction. A few clusters display a limited number of gaps in their evolution (eg, clusters 14, 15, 20 and 22), mostly in correspondence with the financial crisis.



FIGURE 4 Clusters: dynamical composition (part 1).

(a) Clusters composition of the DBHT clusters obtained by calculating detrended log returns over the entire time window 1997–2012. The number of stocks in each cluster is shown on the *y*-axis, with different shades for different ICB industries. (b) For cluster 18 in (a), we have detected at each time window the corresponding "similar" (according to the hypergeometric test) cluster and plotted the composition in time. Zero size corresponds to no "similar" cluster having been found. When more than one "similar" cluster is found, only data for the largest cluster is plotted. Parts (c)–(f) show the same plots as in part (b), but for persistence clusters 4, 8, 7 and 17 respectively.

Journal of Network Theory in Finance

www.risk.net/journal



FIGURE 5 Clusters: dynamical composition (part 2).

(a) For cluster 1 in part (a) of Figure 4 on the facing page, we detected at each time window the corresponding "similar" (according to the hypergeometric test) cluster and plotted the composition in time. Zero size corresponds to no "similar" cluster having been found. When more than one "similar" cluster is found, only data of the largest cluster is plotted. Parts (b)–(f) show the same plots as in part (a), but for clusters 6, 20, 14, 22 and 15 respectively. Colors refer to the legend in Figure 4 on the facing page.

• A few clusters show a persistence in terms of industrial composition as well (this is the case with clusters 4 and, to a lesser extent, 8), but most show a clear evolution. In particular, we can quite clearly distinguish a precrisis state and a postcrisis state; the latter is characterized by a higher degree of mixing of different industries. If, over the precrisis period, we find clusters dominated by one or two industries (technology and industrials in cluster 18, oil and gas in 4 and 15, utilities in 17, consumer services and goods in 14 and 20, financials in 6,

www.risk.net/journal

health care in 22), in the crisis and postcrisis years the industries tend to mix together much more, forming combinations that were not present earlier (oil and gas with basic materials and industrials in clusters 1 and 7, utilities with telecommunications and consumer services in 17, financials with consumer goods and services in 6, health care with utilities and consumer goods in 20). This again shows that the years since the crisis have seen a drop in the reliability of industries as benchmarks to diversify risk.

- Apart from the precrisis and postcrisis dichotomy, in some cases the 2007–8 crisis years show their own features as well. As stated above, some clusters "disappear" during the peak of the crisis (clusters 14, 20 and 22). Many others instead show several peaks in their sizes, together with a sudden increase in the number of industries. This is probably related to the merging of many clusters into fewer, larger clusters during the crisis.
- The cluster containing financial stocks (cluster 6) is worth analyzing further, since it seems to play a role in the outbreak of the financial crisis. Indeed, it shows a clear change in 2007, becoming larger and including an increasing number of different industries (especially health care, technology and consumer services). This pattern is probably connected to the rising importance of the financial industry as a driving factor over the outbreak of the crisis. Interestingly, at the end of 2008, when Lehman Brothers went bankrupt, this cluster suddenly drops to a much lower size (although still higher than precrisis values) and a less mixed composition. This suggests that the financial industry ends up playing a major role in the correlation structure from 2009 onward.

4 DISCUSSION

In this paper, we have investigated the dynamical evolution and nonstationarity of market-correlation structure by means of filtered-correlation networks. In particular, we have focused on PMFGs and the clustering that its topology naturally provides by means of the DBHT method. We have measured the persistence of correlation structure by calculating similarity among clusterings in different time windows, using the adjusted Rand index to quantify the similarity.

Our analyses reveal that the outbreak of the 2007–8 financial crisis marked a transition from relatively high levels of persistence to a much more unstable and changeable structure. The minimum persistence was reached at the end of 2008 when the crisis had fully unfolded. But the decay in persistence had already started in late 2006, well before other warning signs were detectable. Correlation structure persistence eventually recovered in the second half of 2009 with relatively high values until the end of 2011. However, such a persistent structure had distinct features from the precrisis structure, including lower relations with the industrial sector's activities. Notably, since the end of 2011 we have been observing a new decay in persistence, which is signalling another unfolding change in the market structure. This also points out that since 2007 correlation matrixes from historical data, both filtered and unfiltered, have become more unstable and therefore less reliable instruments for risk diversification. Moreover, the decrease in the similarity between correlation-based clustering and the industrial sector implies that portfolio-diversification strategies based on economic activity considerations are expected to become less effective. Furthermore, the analysis of the evolving industrial sector composition of each single cluster reveals that most of them display a clear change with the crisis, which makes them more heterogeneous overall in terms of industrial sectors. In particular, we observed that one cluster, mainly made of financial stocks, experienced a sharp rise in its size and heterogeneity that likely reflected the breakdown of the late-2007 financial crisis. This could give interesting insights in terms of early warning signals that we plan to investigate further in future work.

We also plan to carry out the analyses discussed in this work by using alternative community detection methods on graphs (Fortunato 2010). The comparison of different algorithms is a hot topic in network theory (Aldecoa and Marín 2013; Lancichinetti and Fortunato 2009), and these analyses could give other insights into this issue from the perspective of financial data.

APPENDIX A. ADJUSTED RAND INDEX

Following the notation of Wagner and Wagner (2007), let us call X the set of N objects. Y is a partition into communities of X or simply a clustering: that is, "a set $Y = \{Y_1, \ldots, Y_k\}$ of nonempty disjoint subsets of X such that their union equals X" (Wagner and Wagner 2007). Let us also say we have another clustering Y'. We call the matrix $M = \{m_{ij}\}$ the "contingency table", where

$$m_{ij} \equiv |Y_i \cap Y'_j|,\tag{A.1}$$

ie, the number of objects in the intersection of clusters Y_i and Y'_j . Let *a* be the number of pairs of objects that are in the same cluster in both *Y* and *Y'*, and let *b* be the number of pairs that are in two different clusters in both *Y* and *Y'*. Then, the Rand index is defined as the sum of *a* and *b*, normalized by the total number of pairs in *X*:

$$\mathcal{R}(Y,Y') \equiv \frac{2(a+b)}{N(N-1)} = \sum_{i=1}^{k} \sum_{j=1}^{l} \binom{m_{ij}}{2}.$$
 (A.2)

We can assume a generalized hypergeometric distribution to be the null hypothesis associated with two independent clusterings; we describe this in detail in Appendix B.

www.risk.net/journal

The adjusted Rand index is defined as the difference between the measured Rand index and its mean value under the null hypothesis, normalized by the maximum that this difference can reach:

$$\mathcal{R}_{\rm adj}(Y,Y') \equiv \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} \binom{m_{ij}}{2} - t_3}{\frac{1}{2}(t_1 + t_2) - t_3},\tag{A.3}$$

where

$$t_1 = \sum_{i}^{k} \binom{|Y_i|}{2}, \qquad t_2 = \sum_{j}^{l} \binom{|Y'_j|}{2}, \qquad t_3 = \frac{2t_1t_2}{N(N-1)}.$$
 (A.4)

It turns out that $\mathcal{R}_{adj} \in [-1, 1]$, with 1 corresponding to the case of identical clusterings and 0 to two completely uncorrelated clusterings. Negative values instead show anticorrelation between Y and Y' (that is, the number of pairs classified in the same way by Y and Y' is even less than was expected assuming a random overlapping between the two clusterings).

APPENDIX B. HYPERGEOMETRIC TEST

Following the notation used in Section 3.2, let us call Y_i a generic cluster belonging to the clustering calculated over the entire time window. Let Y'_j be the cluster from clustering X_{T_k} in time window T_k with which we want to compare Y_i in order to find if the number of stocks belonging to both Y_i and Y'_j is sensitively higher than was expected by a random overlapping. This can be translated into a statistical one-tail hypothesis test, in which the null hypothesis is the hypergeometric distribution. Say k is the number of stocks Y_i and Y'_j have in common, whereas $|Y_i|$, $|Y'_j|$ are the cardinalities of the two clusters; then, the hypergeometric distribution reads (Feller 2008)

$$P(X = k) = \frac{\binom{|Y'_j|}{k}\binom{N-|Y'_j|}{|Y_i|-k}}{\binom{N}{|Y_i|}}.$$
(B.1)

This distribution is consistent with a scenario in which the overlapping between the two clusters is due purely to chance. For this reason, it is a suitable null hypothesis for testing the similarity between clusters. If P(X = k) so-calculated is less than the significance level, then the test is rejected, and we conclude that the cluster Y_i overexpresses the cluster Y'_j , and they are therefore similar. The significance level of each test performed is 1%, together with the Bonferroni correction for multiple tests (Feller 2008).

Journal of Network Theory in Finance

www.risk.net/journal

DECLARATION OF INTEREST

T. Di Matteo thanks the COST Action TD1210 for partly supporting this work. T. Aste acknowledges the support of the UK Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre (ES/K002309/1).

ACKNOWLEDGEMENTS

The authors thank Bloomberg for providing the data, and the referees for their useful suggestions.

REFERENCES

- Aldecoa, R., and Marín, I. (2013). Exploring the limits of community detection strategies in complex networks. Scientific Reports 3, 2216.
- Anderberg, M. R. (1973). Cluster Analysis for Applications. Academic Press.
- Arthur, D., and Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 1027–1035. SIAM, Philadelphia, PA.
- Aste, T. (2012). An algorithm to compute Planar Maximally Filtered Graphs (PMFGs). URL: www.mathworks.com/matlabcentral/fileexchange/38689-pmfg.
- Aste, T. (2014). An algorithm to compute Directed Bubble Hiererachical Tree (DBHT) clustering. URL: www.mathworks.com/matlabcentral/fileexchange/46750-dbht.
- Aste, T., Di Matteo, T., and Hyde, S. T. (2005). Complex networks on hyperbolic surfaces. *Physica* A **346**, 20–26.
- Aste, T., Shaw, W., and Di Matteo, T. (2010). Correlation structure and dynamics in volatile markets. *New Journal of Physics* 12, 085009.
- Aste, T., Grammatica, R., and Di Matteo, T. (2012). Exploring complex networks via topological embedding on surfaces. *Physical Review* E **86**, 036109.
- Bartolozzi, M., Mellen, C., Di Matteo, T., and Aste, T. (2007). Multiscale correlations in different futures markets. *European Physical Journal* B 58, 207–220.
- Borghesi, C., Marsili, M., and Miccichè, S. (2007). Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode. *Physical Review* E **76**, 026104.
- Buccheri, G., Marmi, S., and Mantegna, R. N. (2013). Evolution of correlation structure of industrial indices of US equity markets. *Physical Review* E 88, 012806.
- Chetalova, D., Schäfer, R., and Guhr, T. (2014). Zooming into market states. Preprint, arXiv:1406.5386v1 [q-fin.ST].
- Coronnello, C., Tumminello, M., Lillo, F., Miccichè, S., and Mantegna, R. N. (2011). Sector identification in a set of stock return time series traded at the London Stock Exchange. *Acta Physica Polonica* **36**, 2653–2680.
- Di Matteo, T., and Aste, T. (2002). How does the eurodollars interest rate behave? *Journal* of *Theoretical and Applied Finance* 5, 122–127.

www.risk.net/journal

- Di Matteo, T., Aste, T., and Mantegna, R. N. (2004). An interest rate cluster analysis. *Physica* A **339**, 181–188.
- Di Matteo, T., Aste, T., Hyde, S. T., and Ramsden, S. (2005). Interest rates hierarchical structure. *Physica* A **335**, 21–33.
- Di Matteo, T., Pozzi, F., and Aste, T. (2010). The use of dynamical networks to detect the hierarchical organization of financial market sectors. *European Physical Journal* B **73**, 3–11.
- Donetti, L., and Muñoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment* **2004**(10), P10012.
- Fama, E. F., and French, K. R. (2004). The capital asset pricing model: theory and evidence. *Journal of Economic Perspectives* **18**(3), 25–46.
- Feller, W. (2008). An Introduction to Probability Theory and Its Applications, Volume 2. Wiley.
- Fenn, D. J., Porter, M. A., Mucha, P. J., McDonald, M., Williams, S., Johnson, N. F., and Jones, N. S. (2012). Dynamical clustering of exchange rates. *Quantitative Finance* 12(10), 1493–1520.
- Fortunato, S. (2010). Community detection in graphs. Physics Reports 486, 75–174.
- Guimerà, R., Sales-Pardo, M., and Amaral, L. A. N. (2004). Modularity from fluctuations in random graphs and complex networks. *Physical Review* E **70**(2), 025101.
- Hu, Y., Li, M., Zhang, P., Fan, Y., and Di, Z. (2008). Community detection by signaling on complex networks. *Physical Review* E **78**(1), 016115.
- Hubert, L., and Arabie, P. (1985). Comparing partitions. *Journal of Classification* 2, 193– 218.
- Jang, W., Lee, J., and Chang, W. (2010). Currency crises and the evolution of foreign exchange market: evidence from minimum spanning tree. *Physica* A **390**(4), 707–718.
- Kaufman, L., and Rousseeuw, P. J. (1987). Clustering by means of medoids. In *Statistical Data Analysis Based on the L*₁*-Norm and Related Methods*, Dodge, Y. (ed), pp. 405–416. North-Holland.
- Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., and Ben-Jacob, E. (2010). Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS ONE* 5(12), e15032.
- Kohonen, T., Schroeder, M. R., and Huang, T. S. (2001). Self-Organizing Maps, 3rd edn. Springer.
- Lancichinetti, A., and Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical Review* E **80**(5), 056117.
- Livan, G., Inoue, J., and Scalas, E. (2012). On the non-stationarity of financial time series: impact on optimal portfolio selection. *Journal of Statistical Mechanics: Theory and Experiment* **2012**(7), P07025.
- Mantegna, R. N. (1999). Hierarchical structure in financial markets. *European Physical Journal* B 11, 193–197.
- McDonald, M., Suleman, O., Williams, S., Howison, S., and Johnson, N. F. (2008). Impact of unexpected events, shocking news, and rumors on foreign exchange market dynamics. *Physical Review* E **77**, 046110.
- Mitrović, M., and Tadić, B. (2009). Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities. *Physical Review* E **80**(2), 026123.

- Munnix, M. C., Shimada, T., Schäfer, R., Leyvraz, F., Seligman, T. H., Guhr, T., and Stanley, H. E. (2012). Identifying states of a financial market. *Scientific Reports* **2**, 644.
- Musmeci, N., Aste, T., and Di Matteo, T. (2014). Relation between financial market structure and the real economy: comparison between clustering methods. Working Paper, arXiv:1406.0496v1 [q-fin.ST].
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review* E **69**(6), 066133.
- Newman, M. E. J., and Girvan, M. (2004). Community detection in graphs. *Physical Review* E **69**(2), 026113.
- Onnela, J. P., Chakraborti, A., Kaski, K., and Kertész, J. (2003a). Dynamic asset trees and Black Monday. *Physica* A **324**, 247–252.
- Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., and Kanto, A. (2003b). Dynamics of market correlations: taxonomy and portfolio analysis. *Physical Review* E 68, 056110.
- Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., and Kanto, A. (2003c). Asset trees and asset graphs in financial markets. *Physica Scripta* **T106**, 48–54.
- Pozzi, F., Di Matteo, T., and Aste, T. (2012). Exponential smoothing weighted correlations. *European Physical Journal B* **85**(6), 175, 295.
- Pozzi, F., Di Matteo, T., and Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific Reports* **3**, 1665.
- Reichardt, J., and White, D. R. (2007). Role models for complex networks. *European Physical Journal B* 60, 217–224.
- Rosvall, M., and Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7327–7331.
- Ruan, J., Dean, A., and Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Systems Biology* **4**(8), 1–21.
- Shi, J., and Malik, J. (2000). Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888–905.
- Song, W. M., Di Matteo, T., and Aste, T. (2011). Nested hierarchies in planar graphs. *Discrete Applied Mathematics* **159**, 2135–2146.
- Song, W. M., Di Matteo, T., and Aste, T. (2012). Hierarchical information clustering by means of topologically embedded graphs. *PLoS ONE* **7**(3), e31929.
- Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control* 32(1), 235–258.
- Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10421–10426.
- Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior and Organisation* **75**, 40–58.
- Tumminello, M., Miccichè, S., Lillo, F., Piilo, J., and Mantegna, R. N. (2011a). Statistically validated networks in bipartite complex systems. *PLoS ONE* 6(3), e17994.
- Tumminello, M., Miccichè, S., Lillo, F., Varho, J., Piilo, J., and Mantegna, R. N. (2011b). Community characterization of heterogeneous complex systems. *Journal of Statistical Mechanics: Theory and Experiment* 1, P01019.

www.risk.net/journal

Wagner, S., and Wagner, D. (2007). Comparing clusterings: an overview. Technical Report, ITI Wagner, Faculty of Informatics, Universität Karlsruhe (TH).

West, D. B. (1996). Introduction to Graph Theory. Prentice-Hall, Englewood Cliffs, NJ.

Zhou, H. (2003). Distance, dissimilarity index, and network community structure. *Physical Review* E **67**(6), 061901.

Journal of Network Theory in Finance

www.risk.net/journal



Research Paper

A multiplex network analysis of the Mexican banking system: link persistence, overlap and waiting times

José-Luis Molina-Borboa¹, Serafin Martínez-Jaramillo¹, Fabrizio López-Gallo¹ and Marco van der Leij^{2,3,4}

¹Directorate General of Financial Stability, Banco de México, Ave. 5 de Mayo 2, Mexico City, Distrito Federal, 06059 Mexico; emails: jmolina@banxico.org.mx, smartin@banxico.org.mx, flopezd@banxico.org.mx

²Center for Nonlinear Dynamics in Economics and Finance, CeNDEF University of Amsterdam, Valckeniersstraat 65–67, 1018 XE Amsterdam, The Netherlands; email: M.J.vanderLeij@uva.nl
 ³Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands
 ⁴De Nederlandsche Bank, Research Department, Westeinde 1, 1017 ZN Amsterdam, The Netherlands

(Received October 1, 2014; revised December 19, 2014; accepted February 7, 2015)

ABSTRACT

This paper analyzes the persistence and overlap of relationships between banks in a multiplex decomposition of the exposures network. Our analysis may be useful for researchers designing stress tests or models in which the behavior of banks is modeled explicitly. This has not been looked at previously, considering the time period involved and the different types of exposures and interactions used. We show that trading relationships overlap for some pairs of banks, and link persistence is higher in the secured than the unsecured market. Moreover, link persistence in the securities cross-holding network is much higher than in other funding networks, and overlap with the other segments of activity is low, despite being persistent over time. Additionally, unsecured loans received by large banks have the shortest waiting times (that is, for a given borrower and a given lender, the number of days elapsed before a

Corresponding author: José-Luis Molina-Borboa Print ISSN 2055-7795 | Online ISSN 2055-7809 Copyright © 2015 Incisive Risk Information (IP) Limited

100 J.-L. Molina-Borboa et al

new loan is observed) regardless of counterparty size, which suggests quicker access to liquidity. Large banks lend (unsecured) with shorter waiting times to medium-sized banks than to small banks; this is not the case when they lend in the secured layer of the network. Small banks have quicker access to liquidity in the secured lending layer when borrowing from medium-sized banks.

Keywords: financial networks; multiplex networks; link persistence; link overlap; waiting times.

1 INTRODUCTION

The study of financial systems, which are inherently complex, by means of network models has increased recently, benefiting from a large toolbox of quantitative, algorithmic and theoretical results from the well-established fields of social and complex networks. Moreover, recent financial crises have shown how interconnected the global financial system is. Additionally, financial institutions interact in different markets, which can be thought of as different layers of networks. This gives rise to a rich set of complex interactions among these layers, each with different topological properties.

Unfortunately, in financial networks it is common to aggregate all of the available data (for example, all of the different types of exposures among financial institutions) into a static weighted directed graph and study only the resulting single structure. Despite its attractiveness, this approach is far too simplistic, and some types of analysis cannot be made if the different dynamics of each interaction layer are ignored. While modeling contagion using each layer independently can lead to an underestimation of systemic risk, as shown in Poledna *et al* (2014), insightful information regarding interactions is lost when the multiplex structure is neglected.

In our research, we build networks based on the daily interactions between banks across different markets in order to investigate the persistence and the overlap of trading relationships. The usefulness of network similarity measures in finding persistences over time is also studied in this paper. By analyzing different layers, we can identify important relationships between banks and discover their tendency to prevail even in times of stress, which may not be evident when only one network at a time is analyzed. The method sheds light on interactions between different layers in the multiplex structure of financial networks. The aims of this paper are several. First, given that much of the empirical evidence on the structure of financial and banking networks has been done at the aggregated level of the exposures (or by considering only the unsecured lending layer), we want to provide empirical evidence on the differences of the emerging structures of interactions between banks.¹ Second, we

¹This is very important when modeling decisions regarding bank behavior in stress testing and systemic risk studies are being faced.
want to quantify the degree of overlap between layers as well as the level of persistence of the relationships in different markets in order to determine the degree of homogeneity/heterogeneity in these two aspects of interbank trading relationships. Studying the waiting times between loans contributes to this paper's goals.²

Another important research question arises: can persistence and overlap be explained by the involved banks' characteristics (financial or topological)? In the last part of the paper, we approach this question and obtain the following answers.

- First, unsecured loans received by large banks have the shortest waiting times, regardless of the size of the lending bank. This suggests that they have the quickest access to the system's liquidity. Also, when large banks lend repeatedly in the unsecured market, they do so most often to large banks. The waiting time increases as the receiving bank's size decreases (lending with shorter waiting times to medium-sized banks than to small banks): this is no longer observed when the banks lend in the secured lending layer. This may be related to the fact that big banks sometimes have larger liquidity needs than small banks. Another possible explanation is that the collateral involved seems to account for the risk associated with lending to smaller banks. The only case in which small banks have quicker access to new loans is in repurchase agreements (repos) made by medium-sized banks, which is where they seem to turn when they are in need of more liquidity.
- Second, in the unsecured market, smaller banks tend to lend to larger banks more than the other way around. However, this does not appear to happen in the secured market. It seems that the collateral involved in the loan makes it easier for larger banks to lend to smaller ones. This can be seen more clearly when a five-day time window is used because more relations become stable in this case (meaning that they are present at least once every five days, rather than daily). Something that can be appreciated when a daily window is used is that the most stable links in both markets come from loans made by banks of smaller sizes.

Through the investigation of whether relationships are persistent, more suitable models for bank reactions under stress can be built. For example, in Fourel *et al* (2013) the authors design a mechanism to decide which links banks will sever (and in which segment of the market) following a shock. This could be based on empirical evidence instead of a rule of thumb; the overlap across layers can be used to determine the possible reactions by banks under stressful circumstances. To illustrate this point, and considering that the overlap and persistence in the cross-holding securities layer

 $^{^{2}}$ For a given borrower and a given lender, the waiting times of a loan are obtained by counting the number of days that have elapsed between new loans (this is further explained in Section 4).

are very high, we assume banks are more likely to sever links in the unsecured lending market (especially those links that show weak persistence) than sell the securities of other banks.

One of our most important long-term goals along this line of research is to be able to calibrate link prediction models in the context of financial networks. Developing reliable link prediction models will allow us to derive the most likely network under different economic conditions by considering the rich database of interactions among banks. This is of particular importance in stress testing.

This paper contributes to the literature by analyzing the persistence and stability of links as well as their overlap across different layers. We also show that the study of multiplex network structures can unveil structures and relationships that cannot be seen with aggregated data. Likewise, we compare the dynamics and structures of the different layers and determine whether the interconnectedness between some banks is present in more than one layer. Our research aids in understanding the dynamics and motives behind banks' interactions.

In addition, we investigate whether the most interconnected banks in one market are the same in the other markets. The possibility of identifying relevant (highly interconnected) players in different markets beyond simple aggregated measures is an important task for financial authorities, given the recent evidence that the financial system is highly interconnected. Nevertheless, not much work has been done on assessing the role of the multiplex structure in the interconnectedness of a system. We do so by studying the evolution and composition of the core in different layers of the banking system.

Our research benefits from the availability of daily interbank matrixes of exposures and transactions. Nevertheless, an important aspect we deal with is the time dimension when measuring link persistence and overlap; it would be unfair, for example, to judge link persistence to be weak for a given pair of banks if they interact three out of five days simply because the dates were not consecutive. For this reason, we resort to a time-aggregation process, aggregating the different layers of the networks with a five-day time window, and investigate the impact on the measurement of link persistence and overlap. This is also investigated by Finger *et al* (2013), who state that aggregating networks in time allows us to discover the representative characteristics of the underlying "latent" network. We rely on their work to highlight the importance of the time aggregation process.

The use of a "sensible" aggregation period should ensure that we extract stable features (if they exist) of the banking network rather than noisy trading patterns at different points in time. In this regard, it is important to investigate the stability of the link structure in order to assess whether subsequent occurrences of the network share many common links.

Finger et al (2013)

Journal of Network Theory in Finance

Multiplex networks are systems in which the nodes interact, sometimes repeatedly, under different layers. The most important characteristic of these networks, in contrast with the simple collection of individual layers, is that relevant aspects might be ignored if each individual network is treated in an isolated way.

Many of the world's natural and man-made networks possess a multiplex structure, as stated in Wasserman and Faust (1994), Cardillo *et al* (2013) and Menichetti *et al* (2014). The multiplex approach to studying different types of networks was introduced in sociology (Coleman 1988; Verbrugge 1979) and engineering (Chang *et al* 1996; Little 2002). More recently, this paradigm has benefited from specific theoretical treatment (De Domenico *et al* 2013b; Leicht and D'Souza 2009) and been employed in many different contexts, such as social networks (Szell *et al* 2010), epidemic processes (Granell *et al* 2014), random networks (Domenico *et al* 2013c) and evolutionary game theory (Gómez-Gardeñes *et al* 2012).

Financial networks, in the context of financial contagion and systemic risk, have also benefited from the use of the multiplex approach (see Bargigli *et al* 2013; Iori *et al* 2014; Kok and Montagna 2013; León *et al* 2014; Miranda and Tabak 2013; Poledna and Thurner 2014; Poledna *et al* 2014; Squartini and Garlaschelli 2014; Thurner and Poledna 2013).

A look at different networks can provide us with a wider interpretation of interconnectedness, and many concepts in networks can be reformulated in the multiplex context. This has been done with the clustering coefficient in Cozzo *et al* (2013), with the concept of centrality in Solá *et al* (2013) and with other structural measures in Battiston *et al* (2014). Reciprocity can also be reformulated under a multiplex framework, which allows institutions to interact reciprocally across different markets, depending on their own profiles and strategies. Reciprocity plays an important role in identifying the persistence of relationships in a banking system.³

The rest of this paper is organized as follows. Section 2 describes the data used in the study, while Section 3 presents the evolution and compares the different layers studied in this paper. Section 4 explains the concept of waiting times and includes interesting results. Section 5 introduces the concepts used to measure the persistence and stability of relationships, while Section 6 presents the tools used to measure overlap. Section 7 discusses the need for an aggregation process from the daily data. Finally, Section 8 closes the paper with the most important results and Section 9 concludes.

www.risk.net/journal

³ Reciprocity is calculated as the ratio of bidirectional links to the total number of linked pairs of nodes.

2 DATA

Through regulatory reports, the Mexican Central Bank can compile complete information on interbank loans, both secured and unsecured, on a daily basis. Further, given the many different transactions reported by banks, it is possible to compute the daily cross-holding of securities between banks.

Our period of study covers January 2007 to September 2013, unless otherwise indicated. The only financial institutions included in this study are banks, due to the broad nature of available information and the fact that the banking system is still the most important part of the Mexican financial system. In all, forty-three institutions are considered. There were thirty active institutions at the beginning of the time period. This number grew to thirty-seven by the end of the first year, and over forty banks have been present since April 2008. Several institutions joined the network, two banks left the system and two banks merged toward the end of the time period. All banks established in Mexico are covered in this study.⁴

The following layers of the network are analyzed.

- Repos: collateralized loans between banks (total amount).
- NDL (new deposits and loans): uncollateralized transactions agreed on the current day (total amount).
- Securities: exposures arising from cross-holdings of other banks' securities.
- ODL (outstanding deposits and loans): just like NDL but including all deposits and loans outstanding on the current day.
- Derivatives: exposures arising from derivatives contracts between banks.

The following pairs of network layers are analyzed due to practical interpretation whenever there is an overlap.

- NDL and securities: a link in both these layers occurs when a lending bank, which already owns securities of the borrowing bank, lends new money to the borrowing bank through an uncollateralized loan.
- NDL and repos: this is perhaps the most important combination, indicating that a bank lent money to another through both a collateralized loan and an uncollateralized loan.

⁴ In Mexico, subsidiaries of foreign banks (which, in fact, account for the majority of the system's assets) exist, rather than branches.

- Securities and ODL: if one bank has a link to another in these two layers, this indicates that the bank is exposed to the other bank in two ways: by acquiring (or rolling over) debt from the borrowing bank and granting an outstanding (or new) uncollateralized loan.
- Securities and repos: a link in both these layers occurs when a lending bank, which already owns securities of the borrowing bank, lends new money to the borrowing bank through a collateralized loan.
- ODL and repos: this indicates that a bank lent money to another through a collateralized loan while maintaining an outstanding exposure with the borrower through an uncollateralized loan.

The unsecured interbank market is the ultimate market for liquidity. In Mexico, this market plays an important role in leveling out liquidity between banks at the end of the business day. However, the repo market is the most important source of funding for banks; it should be noted that the interbank repo market represents only a fraction of the total funding (most of which comes from legal entities and individuals). However, trading relationships are more important in the unsecured interbank market, since there is no collateral involved in this type of loan.

The cross-holding of securities is an interesting layer of the interbank network because, unlike the other layers mentioned, this network does not involve bilateral transactions. However, this layer represents an important funding source for many banks. Persistence in this layer implies a rollover of the securities, meaning that the lending bank "trusts" the borrowing bank.

3 EVOLUTION AND COMPARISON BETWEEN LAYERS

As previously noted, it is extremely important to consider the complete network of interbank exposures in order to accurately measure systemic risk and the way monetary losses could propagate between institutions in a financial system. Attempting to model contagion risk by assessing each of the layers independently can lead to a substantial underestimation of the outcome of an external shock (Poledna *et al* 2014). However, insight and information regarding the way banks interact in a financial system can be lost when this multilayer structure is neglected.

When considering the growth or evolution of the interbank network, studying each layer separately allows us to understand whether changes are due to the common evolution of all layers or the marginal changes present in a subset of the layers.

We observed an increase in the density of the aggregated network. In Figure 1 on the next page, it can be seen that the derivatives network shows the highest volatility among all the layers, while the securities network has grown more than the others: it



FIGURE 1 For each date, the evolution of the density and the total volume are shown for each of the layers.

(a) Daily densities of the different layers. (b) Daily volumes of the different layers (billions of MXN).

has more than doubled its density during the complete time period. The total volume has also increased significantly for the interbank repurchase agreements.⁵

Consider now the reciprocity of interactions between institutions (Figure 2 on the facing page), which can potentially explain the stability of links in a network.⁶ As before, it is clear that different layers do not exhibit the same behavior.

Accordingly, after unveiling the differences between the layers, we can reveal interactions between banks that are persistent through time and may overlap across different markets; thus, it is vital to consider the multiplex structure of the interbank network. In this way, a better understanding of how banks interact can be achieved.

3.1 Resemblance of core-periphery structures across layers

In addition to analyzing the evolution of the different layers of the network, we seek to determine whether the more central banks are the same in the aggregated network than in each of the layers. The cores of the networks were obtained using the methodology introduced in Craig and von Peter (2010) and largely explored in van Lelyveld and

⁵ The density of a network is calculated as the number of links observed divided by the total number of possible links (that is, L/n(n-1), where L is the number of links observed in a directed network with n nodes).

⁶ Reciprocity is calculated as the ratio of bidirectional links to the total number of linked pairs of nodes.



FIGURE 2 Evolution of the fraction of reciprocal links in each layer.

in 't Veld (2012), Martínez-Jaramillo *et al* (2014a) and Langfield *et al* (2014).⁷ We borrow an intuitive definition provided in Craig and von Peter (2010) for the core and the periphery. The core is a subset of intermediaries, excluding those that do not play an essential role in holding the interbank market together. The nodes excluded from the core are in the periphery. For a more precise algorithmic description, we refer the reader to Craig and von Peter (2010).

As with the evolution of the densities, it can be seen in Figure 3 on the next page that a couple of layers behave differently to the rest in the evolution of the sizes of their cores.

Moreover, the banks that belong to the cores of the different layers are not always the same. In Figure 4 on page 109, while there seems to be a couple of banks consistently belonging to most of the cores, the number of banks belonging to one and/or two cores is not stable and varies through time, suggesting that banks that are not important in one layer could be considered as such in another. This information is lost when only the complete network is considered.

In Figure 5 on page 110, it can be observed that only a few banks belong to the core in two layers on the same day. In NDL versus repos (part (ii)), it is interesting to note that only two banks consistently belong to the cores of the networks of new loans, both unsecured and secured; a third bank has, though, become more important recently.

www.risk.net/journal

⁷ We would like to thank Ben Craig and Goetz von Peter for providing us with the code to fit the core–periphery model to the Mexican interbank networks.



FIGURE 3 Evolution of the core sizes of the different layers.

4 WAITING TIMES BETWEEN NEW LOANS: SECURED VERSUS UNSECURED

In Cocco *et al* (2009), an approach is introduced to estimate the probability that a lender who has lent to a borrower on day t_d will do so again on the next k days, t_{d+1}, \ldots, t_{d+k} . Cocco *et al* also compute these probabilities separately for loans going to and from banks of different sizes.

In this section, we adopt the same idea but instead estimate the waiting times for a new loan in the repos and NDL layers (the layers that represent daily new loans) in an attempt to find differences between the collateralized and uncollateralized markets.

For each pair of banks (and each of the mentioned layers), we compute the following vector, which contains the days in which a loan was made from bank i to bank j:

$$Y_{ij}^{G} = \{m \mid a_{ij}^{G}(t_m) = 1, m = 1, 2, \dots, T\},$$
(4.1)

where $a_{ij}^G(t_m)$ denotes the (i, j) entry of the adjacency matrix A on day t_m for the network corresponding to the market G. Also, m = 1, 2, ..., T, and G is either NDL or repos. Afterwards, we compute the differences between the consecutive elements of each Y_{ij} to obtain the waiting times between loans from i to j in the layer G. We now analyze the distribution of these waiting times.

First, it can be observed in Figure 6 on page 111 that the waiting times for repo transactions (red line) are lower than for new deposits and loans (blue line).⁸ We

⁸A highly significant (*p*-value \approx 0) one-sided Kolmogorov–Smirnov difference between the two empirical distributions.



FIGURE 4 Number of banks in different cores.

separated the banks into three groups, small, medium and large, according to their assets. For each day, the banks were ordered according to their assets' sizes. The ten largest banks were categorized as "big". The remaining banks were split into two groups: the first group was categorized as "medium" and the second one as "small".⁹

Figure 7 on page 112 shows the different behaviors that can be perceived when this distinction is made.

The loans received by large banks are the ones that have the shortest waiting times for NDL, regardless of the size of the lending bank, which means they have the

www.risk.net/journal

⁽a) In no cores. (b) In one core. (c) In two cores. (d) In three cores. (e) In four cores. (f) In all five cores.

⁹ The decision to split the system this way was supported by the fact that ten banks consistently represented a significant part of the system, while fifteen banks were consistently much smaller. This way, it was clear which banks were big and which were small, so the remaining were categorized as medium.



FIGURE 5 Intersection of members between cores.

(a) NDL versus securities. (b) NDL versus repos. (c) Securities versus ODL. (d) Securities versus repos. (e) ODL versus repos.

quickest access to the system's liquidity.¹⁰ From this, it can also be seen that when large banks lend repeatedly in NDL, they do so more often to large banks, and the waiting time increases as the receiving bank's size decreases. It is interesting to note that large banks in NDL lend with shorter waiting times to medium banks than to

¹⁰A highly significant one-sided Kolmogorov–Smirnov distance for all cases, except for the case of medium banks lending to small banks.



FIGURE 6 Empirical cumulative distribution functions for the times between loans for NDL (blue, below) and repos (red, above).

small banks.¹¹ This is no longer observed when they lend in the repos layer, where the collateral involved seems to account for the apparent lack of desire from large banks to lend to smaller banks in the uncollateralized market.¹² The only case in which small banks have quicker access to new loans is in repos made by medium banks, which is where they seem to turn when in need of more liquidity.¹³

It is worth reiterating that the goal of this exercise is to study relationships between banks of different sizes and determine whether the frequency of loans is different depending on the size of the banks. Studying the waiting times between loans is useful in understanding the persistence of relationships in the interbank market. It should be noted, however, that the individual liquidity needs of some banks could sometimes be the main cause of their specific waiting times (for example, small banks may satisfy all their liquidity needs with only one sporadic loan from a big bank, causing a longer waiting time).

4.1 Waiting time to recover loans

In this exercise, we choose a reference network and count the fraction of links of this network that have been observed again after a certain number of days. We look at the fraction of loans repeated after a certain number of newly observed days for NDL (blue) and repos (red).

¹¹ For the Kolmogorov–Smirnov one-sided test, the reported *p*-value was 0.001135.

¹² No significant Kolmogorov–Smirnov distance.

¹³ A highly significant one-sided Kolmogorov–Smirnov distance for both cases.



FIGURE 7 Empirical cumulative distribution functions for the times between loans.

(a) NDL granted by small banks. (b) Repos by small banks. (c) NDL granted by medium banks. (d) Repos by medium banks. (e) NDL granted by large banks. (f) Repos by large banks. The first column corresponds to new deposits and loans, while the second column corresponds to new repo loans. The rows correspond to small, medium and large banks. The color code is as follows: blue represents loans received by large banks, red represents loans received by medium banks and green represents loans received by small banks.

Journal of Network Theory in Finance

FIGURE 8 Fraction of links of the reference network that have been observed again after a certain number of days (in blue, uncollateralized loans; in red and dashed, collateralized loans).



(a) First day of January 2011 and (b) first month of 2011 used as reference network.

In part (a) of Figure 8, the first day of 2011 is chosen as the reference network. It can be seen that most of the links are observed again within a few days, with the loans of the uncollateralized market (blue line) seemingly repeating faster than in the collateralized market (red line).

Part (b) of Figure 8 shows the result using the sum of all networks in January 2011 as a reference network, so that the first day on the *x*-axis corresponds to the first day of February 2011. Obviously, the accumulation is slower than in part (a) of Figure 8, but the speed at which most of the loans are seen again is remarkable. Also, the difference between the uncollateralized (blue) and the collateralized (red) markets is greater in this case.

4.2 Fraction of loans repeated after different periods

In a similar exercise, we calculated, for each day and for both the collateralized and the uncollateralized markets, the fraction of a day's loans that were observed within the next day, the next week, the next month and the next quarter. The results are shown in Figure 9 on the next page.

As previously shown, the collateralized market (part (b) of Figure 9) seems to repeat loans faster. However, an increase during the first half of the time span can be observed for the number of repeated loans in the uncollateralized market (part (a) of Figure 9), especially for the time windows of one day and one week (blue and red lines, respectively).

www.risk.net/journal

FIGURE 9 For each date, each line counts the fraction of observed loans that were observed again within the next day (orange line), the next week (red line), the next month (green line) and the next quarter (blue line).



(a) NDL: uncollateralized market. (b) Repos: collateralized market

For both markets, many of the loans are repeated after one month, and the vast majority are repeated after one quarter.

In conclusion, studying the waiting times across different markets can help us understand the stability of the loans observed in a network in terms of how likely it is that a loan will exist given that it has existed before. The comparison between layers helps us to understand the different dynamics when collateral is present (repos layer) and when it is not present (NDL layer). Last, differentiating the size of the lending banks reveals valuable information with regard to the interaction and reliability of links when the size of the banks involved is taken into account.

5 MEASURING PERSISTENCE AND STABILITY OF RELATIONSHIPS

In this section, we investigate the different layers of the interbank network to find relations between institutions that persist in consecutive natural days. Given that there is a link present for a certain day, we are interested in discovering whether there is a nonnegligible probability that the link will be present the next day (Nicosia *et al* 2013).

We explore three ways of measuring the persistence and stability of links across time: topological overlap, variation and stability matrixes and similarity measures. Remember that $a_{ij}(t_m)$ will denote the (i, j) entry of the adjacency matrix A on day t_m (loans from i to j), where m = 1, 2, ..., T.

5.1 Topological overlap

Following Nicosia *et al* (2013) and Hidalgo and Rodriguez-Sickert (2008), given two networks consecutive in time, the topological overlap of the neighborhood of *i* in $[t_m, t_{m+1}]$ is defined as

$$TO_i(A(t_m), A(t_{m+1})) = \frac{\sum_j a_{ij}(t_m)a_{ij}(t_{m+1})}{\sqrt{[\sum_j a_{ij}(t_m)][\sum_j a_{ij}(t_{m+1})]}}.$$
(5.1)

Averaging over all days, we obtain the average topological overlap of *i*'s neighborhood, which can be thought of as the probability that an edge from *i* to one of its neighbors persists across two consecutive days, or as the tendency of *i*'s links to persist across multiple days. If, instead of averaging over all days, we obtain the daily average over all banks' topological overlaps, we obtain a measure of the tendency of all nodes' links to persist across days t_m and t_{m+1} :

$$TO(A(t_m), A(t_{m+1})) = \frac{1}{N} \sum_{i=1}^{N} TO_i(A(t_m), A(t_{m+1})).$$
(5.2)

Figure 10 on the next page shows the series for TO_m^G , where *G* corresponds to one of the following layers: {repos, NDL, securities, ODL}. It can be seen that the topological overlap across consecutive days is highest for the securities layer. Also, it is interesting to note that the repos and NDL layers follow the same dynamics, but repos seems to be consistently higher than NDL. In turn, NDL accounts for a great part of the overlap of ODL, meaning that it is common for banks to make two new loans on two consecutive days rather than a loan with a two-day term. In related works, we have found that overnight loans dominate the unsecured market (Martínez-Jaramillo *et al* 2014b).

5.2 Stability and variation matrixes

The following method is proposed in Bellenzier (2013) to quantify the persistence of the edges in a time interval given different behaviors of a link on two consecutive days. The stability matrix for day t_m , denoted $D(t_m)$ (m = 1, 2, ..., T-1), is defined as follows:

$$d_{ij}(t_m) = \begin{cases} -1 & \text{if } a_{i,j}(t_m) = a_{i,j}(t_{m+1}) = 0, \\ 1 & \text{if } a_{i,j}(t_m) = a_{i,j}(t_{m+1}) = 1, \\ 0 & \text{otherwise.} \end{cases}$$
(5.3)

Bellenzier further proposes introducing new variables to quantify link activity (when links either exist or change their status on two consecutive days), among others. However, we are going to concentrate on using the link stability matrix, S, which

www.risk.net/journal



FIGURE 10 Evolution of the topological overlap of the different layers.

counts the number of consecutive days on which a link persists:

$$s_{ij} = \sum_{t_m = 1 | d_{ij}(t_m) \neq -1}^{T-1} d_{ij}(t_m).$$
(5.4)

The histograms in Figure 11 on the facing page show the (standardized) nonzero values of the *S* matrix for each of the different layers, ie, the fraction of consecutive days for which a link persisted. Confirming what has been claimed before, the securities layer has the fattest tail, meaning that it has the highest fractions of persistent links across two consecutive days. However, comparing it with the ODL layer, while the highest probabilities correspond to securities, the more extreme cases (links that persisted almost all of the time) are seen in the ODL layer. Comparing repos and NDL, collateralized loans persisted more often than uncollateralized loans.

5.3 Similarity measures

A similarity-based approach, using the Jaccard index, was used to quantify the similarity of networks that are one day apart (a Jaccard self-similarity with lag equal to 1). According to Bargigli *et al* (2013), the Jaccard similarity can be interpreted as the probability of observing a link in a network conditional on the observation of the same link in the other network. In a banking system with *n* institutions, having an adjacency matrix in two different time periods, $A(t_m)$ and $A(t_{m+1})$, in which each entry can have the values 0 or 1, we define the following metrics. M_{11} is the total



FIGURE 11 Distribution of the standardized nonzero values of the link stability matrix for different layers.

(a) Repos. (b) NDL. (c) Securities. (d) ODL. (e) Derivatives. The local maximums at the right-hand end of some histograms represent relationships that were always stable.

number of entries for which

$$a_{ij}^{A(t_m)} = a_{ij}^{A(t_m+1)} = 1, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

 M_{01} represents the total number of entries for which $a_{ij}^{A(t_m)} = 0$ and

$$a_{ij}^{A(t_{m+1})} = 1, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

www.risk.net/journal



FIGURE 12 Evolution of the Jaccard self-similarities of the different layers.

 M_{10} represents the total number of entries for which $a_{ij}^{A(t_m)} = 1$ and

$$a_{ij}^{A(t_{m+1})} = 0, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

And finally, M_{00} represents the total number of entries for which $a_{ii}^{A(t_m)} = 0$ and

$$a_{ij}^{A(t_{m+1})} = 0, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

The Jaccard index is equal to the ratio of the number of links present in both matrixes to the total number of links present in any of the two matrixes:

$$J(A(t_m), A(t_{m+1})) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$
(5.5)

The behavior in Figure 12 is similar to that observed in the topological overlaps (Figure 10 on page 116), but the series seem to be more volatile before 2010, with the securities layer being slightly lower for that period. The gap between the persistence of repos and NDL, which was mentioned earlier, becomes more clear using the self-similarity method.

6 OVERLAP OF LINKS ACROSS LAYERS

In this section, we are interested in finding the links present in different layers on the same day. This is important because it may reveal information about relationships between banks that would otherwise be ignored in the aggregated network, given that,

when observing a link in the aggregated network, we cannot tell how many of the layers had the link present.

This section presents the methods (which are merely a modification of those in Section 5) that will be used to study the overlap of links. The analysis is designed to compare networks corresponding to the same day, although we will discuss later how changing the time window (currently equal to one day) can change the results significantly. Results using both time windows will be shown in Section 8.

To evaluate the significance of the results that have the form of time series, comparisons with null models, similar to those in Maslov and Sneppen (2002), were made. For each daily network of each layer, 100 matrixes were simulated by preserving the out-degrees of all nodes and shuffling the endpoints of the nodes. To evaluate the significance of the overlap of two layers, the metrics were calculated for 100 pairs of networks (obtained from the 100 simulations of each of the layers), and observed results (from real data) were then compared with the mean and standard deviation of the "null" metric of overlap.

In conclusion, despite using a very direct method to detect overlap, we found that relationships that occur simultaneously in two or even three layers on the same day exist, which suggests an important funding relationship.

6.1 Overlap

Part (a) of Figure 15 on page 124 and part (a) of Figure 16 on page 126 count the number of links that were present in two or three layers on the same day. For the series pairs, the five combinations introduced earlier are shown in part (a) of Figure 15.

For the entries present in three layers at the same time (shown in part (a) of Figure 16), part (i) counts links that represent lending banks, which, given that they already owned securities of the borrowing bank, funded that same counterparty with both a repo transaction and an NDL, all on the same day. Part (ii) is analogous but considers ODLs (which include NDL as well).

6.2 Topological overlap across layers

As mentioned, a slight modification of the approach introduced in Section 5 was used. Instead of calculating the topological overlap between two networks of the same layer one day apart, we calculated the overlap using two networks corresponding to the same day for two different layers:

$$TO(G(t_m), G'(t_m)) = \frac{1}{N} \sum_{i=1}^{N} TO_i(G(t_m), G'(t_m)),$$
(6.1)

www.risk.net/journal

where $G(t_m) \neq G'(t_m)$ correspond to one of the following layers: {repos, NDL, securities, ODL}. Part (a) of Figure 17 on page 127 shows the evolution of the topological overlap for the five combinations of interest.

6.3 Stability and variation matrixes across layers

Similarly, we obtain stability and variation matrixes across layers by using two networks corresponding to the same day for two different layers. In the end, we obtain similarity matrixes $S^{A,B}$ whose entries $s_{ij}^{A,B}$ count the number of days in which the link (i, j) was present in both layers, A and B.

The distributions of the nonzero values of the link stability matrixes are shown in part (a) of Figure 18 on page 128.

6.4 Similarities across layers

In a similar way to (5.5), we can define the Jaccard index for two different layers at the same period of time, $G(t_m)$ and $G'(t_m)$, in a banking system with *n* banks. Now we have that M_{11} is the total number of entries for which

$$a_{ij}^{G(t_m)} = a_{ij}^{G'(t_m)} = 1, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

 M_{01} represents the total number of entries for which $a_{ii}^{G(t_m)} = 0$ and

$$a_{ij}^{G'(t_m)} = 1, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

 M_{10} represents the total number of entries for which $a_{ij}^{G(t_m)} = 1$ and

$$a_{ij}^{G'(t_m)} = 0, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

And finally, M_{00} represents the total number of entries for which $a_{ij}^{G(t_m)} = 0$ and

$$a_{ij}^{G'(t_m)} = 0, \quad \forall i \neq j \in \{1, 2, \dots, n\}.$$

The Jaccard index for two different layers in the same period of time is calculated as follows:

$$J(G(t_m), G'(t_m)) = \frac{M_{11}}{M_{10} + M_{10} + M_{11}}.$$
(6.2)

The time series shown in part (a) of Figure 19 on page 129 show the results.

Journal of Network Theory in Finance

Layer	OV (%)	TO (%)	SIM (%)
NDL versus repos	46	48	52
NDL versus securities	70	65	28
Repos versus securities	50	43	48
ODL versus repos	70	61	29
ODL versus securities	87	83	12

TABLE 1 The percentage of days in which the observed metric was more than three standard deviations away from that day's null model.

The metrics are the number of overlap links (OV), topological overlap (TO) and Jaccard similarity (SIM).

6.5 Comparison with null models

Table 1 compares the results with our simulations.

7 THE NEED FOR A TIME WINDOW AND ITS CONSEQUENCES

In order to study the overlap between layers, a one-day window may be too harsh due to the high-frequency nature of our data. While in Section 5 we did find evidence of persistence using daily information, the results of Section 6 and the determinants of persistence (both shown in Section 8) suggest that potentially informative evidence is not seen clearly when only daily information is used.

Consider a bank that lends to another bank for fifty consecutive days but alternates between two different markets every day (day 1 in repos, day 2 in NDL, day 3 in repos, day 4 in NDL and so forth). Our one-day window would not capture this overlap between layers, which clearly exists and reveals an important relationship between these two institutions.

While there are a handful of methods to summarize consecutive temporal graphs into a static graph (Nicosia *et al* (2013) and Sharan and Neville (2007) provide a few examples), the choice of an appropriate method and time window for financial data is a difficult one. This is especially true if we are to construct adjacency matrixes (which are the input for most of the topological metrics and require unweighted graphs) to estimate similarity and persistence. If the window is chosen incorrectly, important relationships could be underestimated, while isolated relationships could be deemed significant.

Further, the appropriate time window could be different for different financial systems or even layers of the same financial network. The time window has to be such that, while being economically meaningful, it retains important topological and temporal information. Figure 13 on the next page shows the cumulative density of



FIGURE 13 Densities for cumulative matrixes starting on the first day of 2011.

The blue line represents uncollateralized loans and the red, dashed line represents repo transactions.

the NDL (blue) and repos (red) layers from the first day of 2011 until the end of our data set. $^{\rm 14}$

On the basis of our experience, we believe that it is most adequate to study high-frequency networks by aggregating them along the time dimension. One important paper on the topic of time aggregation in networks is Delpini *et al* (2013). Similar results in terms of changes in the studied metrics were found by Finger *et al* (2013), who noted that

- yearly aggregation shows variation due to changes in the system (such as changes in active banks),
- monthly aggregation is problematic because of monthly seasonality, as shown by Iori *et al* (2008),
- quarterly aggregation seems to be appropriate for this type of study.

The plots of the cumulative densities (Figure 13) for both layers show a concave downward shape, which suggests a decreasing rate in the appearance of new links. However, the sensitivity of the time window is highlighted: a small period of time is considered (less than fifty days, for example) because of the quick increase in density.

In this paper, for demonstrative purposes, a time window of five days will be adopted since it represents one business week. This is small enough to capture shortor medium-term relationships and persistences but large enough to smooth our highfrequency information.

¹⁴ This period was chosen because it was not until 2011 that the number of banks in the system became stable.



FIGURE 14 Reciprocity index for a time window of five days (consider that more one-way links are added as the window grows larger).

7.1 Reciprocity

It is interesting to note the changes observed in reciprocity when considering a fiveday window, as shown in Figure 14. Compared with the daily window used to generate Figure 2 on page 107, the density of outstanding deposits and loans is lower, given the fact that, while there are more days to allow a relationship to become reciprocal, the addition of new (nonreciprocal) loans pulls down the reciprocity index.

In addition, the reciprocities for the unsecured (NDL) and secured (repos) markets are higher using this new five-day window compared with using daily windows. Finally, it has been found that reciprocity is not consistently higher for either of these two layers: there are periods (which last for a few months) in which it is higher for one or the other.

8 RESULTS AND DETERMINANTS OF LINK PERSISTENCE, STABILITY AND OVERLAP

In this section, results will be presented using both the daily and five-day time windows. Additionally, scatterplots coupling link stability and banks' characteristics are included.

8.1 Overlap of links

Figure 15 on the next page and Figure 16 on page 126 show the number of link overlaps on pairs of networks using both time windows.

www.risk.net/journal



FIGURE 15 Evolution of the number of link overlaps between different pairs of layers. (a) One-day time window.

(i) NDL versus securities. (ii) NDL and repos. (iii) Securities versus ODL. (iv) Securities versus repos. (v) ODL versus repos.





(i) NDL versus securities. (ii) NDL and repos. (iii) Securities versus ODL. (iv) Securities versus repos. (v) ODL versus repos.

Journal of Network Theory in Finance

Focusing on part (ii) in parts (a) and part (b) of Figure 15, changing the time window helped us to find what we were looking for: relationships between banks in different layers. When a daily window is used, there seems to be, on average, four to six loans present in both markets, and the time series is significantly larger than when a five-day time window is used. The daily window also reveals important information with regard to the overlap of loans across layers. The same is true for part (i) in parts (a) and (b) of Figure 16, where using a larger window makes it clear that banks exist which own paper from another bank, lend to that bank using a secured loan and lend to that same bank using an unsecured loan – all at the same time.

Comparing part (i) with part (iv) in parts (a) and (b) of Figure 15 on the facing page, the overlap between the cross-holding of securities and the uncollateralized market (NDL) seems to be larger than the overlap with the collateralized market (repos). It also seems to grow more consistently.

8.1.1 Topological overlap across layers

In Figure 17 on page 127, it is interesting to note that, although a slight level increase seems to appear in part (iii), changing the time window does not appear to significantly increase the topological overlap of the lending banks' neighborhoods. While we demonstrated before that increasing the time window helps to detect more overlap between loans, the topological overlap shows that lenders' neighborhoods already overlap enough using a daily window, so increasing it does not add much to the results.

8.1.2 Stability and variation matrixes across layers

Figure 18 on page 128 reveals interesting results and different stories depending on the time window used. First, comparisons between pairs of layers in part (a) of Figure 18 on page 128 show that there is low stability in the overlap of NDL versus repos. Also, if a lending bank holds paper from a borrowing bank, it is more likely to lend to it on the same day through an uncollateralized loan (part (iv), NDL versus securities) than through a collateralized loan (part (v), repos versus securities).

Second, increasing the time window shows that a large number of loans is present in different pairs of layers in over 95% of the weeks observed. While using a daily window does not allow us to identify these overlap loans, the new time window reveals a true and consistent overlap across the complete time period.

From this result, we can understand the relevance of a temporal aggregation process when unveiling persistence and overlap.

8.1.3 Similarities across layers

Figure 19 on page 129 shows that, in general, similarities between layers change greatly over time, even when a larger time window is considered. An interesting fact





(i) NDL/securities/repos. (ii) ODL/securities/repos.

FIGURE 16 (b) Five-day time window.



(i) NDL/securities/repos. (ii) ODL/securities/repos.

is that the similarity measure seems to be decreasing for almost all pairs of layers. In addition, the similarity in NDL versus securities (part (i) of parts (a) and (b)) seems to remain stable for most of the time period, whereas the similarity in NDL versus repos has consistently decreased over time.



FIGURE 17 Evolution of the topological overlaps between different pairs of layers. (a) Oneday time window.

(i) ODL versus repos. (ii) ODL versus securities. (iii) NDL versus repos. (iv) NDL versus securities. (v) Repos versus securities.





(i) ODL versus repos. (ii) ODL versus securities. (iii) NDL versus repos. (iv) NDL versus securities. (v) Repos versus securities.

www.risk.net/journal



FIGURE 18 Distribution of the standardized nonzero values of the link stability matrixes for different layer combinations. (a) One-day time window.

(i) ODL versus repos. (ii) ODL versus securities. (iii) NDL versus repos. (iv) NDL versus securities. (v) Repos versus securities.





(i) ODL versus repos. (ii) ODL versus securities. (iii) NDL versus repos. (iv) NDL versus securities. (v) Repos versus securities. The local maximums at the right-hand end of some histograms represent relationships that were always stable for each time window.

Journal of Network Theory in Finance



FIGURE 19 Evolution of the Jaccard index between different pairs of layers. (a) One-day time window.

(i) NDL versus securities. (ii) NDL versus repos. (iii) Securities versus ODL. (iv) Securities versus repos. (v) ODL versus repos.





(i) NDL versus securities. (ii) NDL versus repos. (iii) Securities versus ODL. (iv) Securities versus repos. (v) ODL versus repos.

www.risk.net/journal

8.2 Relation of stability with other variables

This subsection compares the stability of links (using the entries of the stability matrix s_{ii}) with different average characteristics of the banks across the time period.

Figure 20 on the facing page compares the entries of the stability matrix with the involved institutions' assets for (i) ODL, (ii) NDL and (iii) repos for both windows.

An interesting fact, which can already be seen in part (a) of Figure 20 on the facing page but becomes clearer in part (b) of Figure 20, is that, in the unsecured market (NDL), smaller banks tend to lend to larger banks more than the other way around (in part (i) and part (ii), there are more points below the horizontal axis than above). However, in the secured market (part (iii), repos) this does not appear to happen, as if the collateral involved in the loan makes it easier for larger banks to lend to smaller ones. Once again, this is more clearly seen when using a five-day time window, as more relationships become stable in this case (because they are present at least once every five days, rather than daily). Something that can be appreciated with a daily window is that the most stable links in both markets came from loans made by banks of smaller sizes.

Figure 21 on page 133 shows the relation to clustering coefficients in derivatives for both time windows. Although other works have found a relationship between persistence and reciprocity (such as Hidalgo and Rodriguez-Sickert (2008) or Clauset and Eagle (2012)), not much can really be said in our case, other than that banks with similar coefficients tend to have more stable relationships (although they are by no means the only ones).

Figure 22 shows the relationship of a link's stability with its reciprocity. For some markets, these two variables seem to be related. It is interesting to note that, whenever there is a relationship between these variables, it is more common for stability than reciprocity to be higher, especially when the time window is increased. This is due to the fact that stability grows faster with time than reciprocity. This means that, as the days go by, it is more likely for a loan to be repeated than to become reciprocal.

9 CONCLUSIONS AND FURTHER WORK

While it is impossible to properly evaluate the likelihood of direct contagion in a banking system without a comprehensive concept of exposure, it is also impossible to understand and quantify the strength and persistence of bilateral relationships without first considering the multiplex nature of the financial system.

In this research, despite our stringent requirements to measure link persistence and overlap with daily data, we found evidence of link persistence and overlap between layers. Additionally, with a simple time aggregation study the evidence became clear:



FIGURE 20 Relation of stability to borrower's and lender's asset sizes (fraction of the system's assets). (a) One-day time window.

(i) ODL. (ii) NDL. (iii) Repos.

the layer in which persistence and stability were most easily identified was the securities cross-holding network. This finding is relevant, as not many previous papers have been able to identify such exposures or study this particular layer using such a perspective. As mentioned, this fact can explain why banks that retain other banks' securities show trust in the recipient banks' health without exhibiting drastic changes in confidence.

www.risk.net/journal





(i) ODL. (ii) NDL. (iii) Repos.

Little correlation was found between the compositions of the cores in different layers, implying that the most significant nodes in each market are different. This finding has important implications for supervisors, as it shows that the identification of important players should be executed by taking the multiplex rather than the aggregated network into account.

By analyzing the waiting times between loans, we found that they are shorter for big banks than for medium-sized and small banks in the unsecured market. This phenomenon disappears in the secured lending market. Additionally, large banks lend





Relation of stability to borrower's and lender's clustering coefficients.

with shorter waiting times to medium-sized than to small banks, and small banks resort to medium-sized banks for liquidity in the secured market. At the moment we cannot tell whether this is a supply- or demand-driven phenomenon, given that small, medium-sized and big banks all have different liquidity needs. Further investigation is required here.

One implicit assumption in the study of financial contagion through interbank markets is that the exposures network, once obtained or simulated, is static or at least stable. After collecting all the empirical evidence from this research, we conclude that, by decomposing the full exposures network, we can distinguish different properties across the layers. First, the persistence of the securities cross-holding network is the highest. That is because this layer governs a large part of the persistence for the full exposures network, and banks' reactions are more likely to occur in other layers that possess more dynamic features than the securities cross-holdings network. Second, there are relationships between institutions that cover all the layers in the exposures network. This is very important, because they represent the links that should be preserved in the models that incorporate banks' reactions.

www.risk.net/journal





(i) NDL. (ii) Repos. (iii) Securities. (iv) ODL. (v) Derivatives.





(i) NDL. (ii) Repos. (iii) Securities. (iv) ODL. (v) Derivatives. Relation of stability to the reciprocity of the relation.

Journal of Network Theory in Finance

Among the tasks that remain are the following. First, we need to formally determine a time window for the temporal aggregation process, since we have only started the investigation of the time aggregation process and link persistence. Second, more work needs to be done on the banks' temporal characteristics that determine link persistence and overlap, such as capital ratio, nonperforming loans and return on assets. Finally, we will seek a relationship between the waiting time for a new loan between a pair of banks and their reciprocity index.

DECLARATION OF INTEREST

The opinions expressed here do not reflect the views of Banco de México, the Financial Stability General Directorate or De Nederlandsche Bank.

ACKNOWLEDGEMENTS

The authors are grateful to the audience at the Financial Risk and Network Theory Cambridge Centre for Risk Studies Seminar and to anonymous referees for their useful comments.

REFERENCES

- Afonso, G., Kovner, A., and Schoar, A. (2013). Trading partners in the interbank lending market. Staff Report, Federal Reserve Bank of New York.
- Bargigli, L., di Iasio, G., Infante, L., Lillo, F., and Pierobon, F. (2013). The multiplex structure of interbank networks. Working Paper, Economics, Dipartimento di Scienze per l'Economia e l'Impresa, Università degli Studi di Firenze.
- Battiston, F., Nicosia, V., and Latora, V. (2014). Structural measures for multiplex networks. *Physical Review* E **89**(3), 032804.
- Bellenzier, L. (2013). Dynamics in Financial Networks. PhD thesis, Università degli Studi di Milano-Bicocca.
- Bianconi, G. (2013). Statistical mechanics of multiplex networks: entropy and overlap. *Physical Review* E **87**, 062806.
- Brummitt, C. D., D'Souza, R., and Leicht, E. A. (2012). Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences of the United States of America* **109**(12), E680–E689.
- Cardillo, A., Gómez-Gardeñes, J., Zanin, M., Romance, M., Papo, D., del Pozo, F., and Boccaletti, S. (2013). Emergence of network features from multiplexity. *Scientific Reports* **3**(1344), doi:10.1038/srep01344.
- Chang, S. E., Seligson, H. A., and Eguchi, R. T. (1996). Estimation of the economic impact of multiple lifeline disruption: Memphis light, gas and water division case study. Technical Report NCEER-96-011, University at Buffalo.
- Clauset, A., and Eagle, N. (2012). Persistence and periodicity in a dynamic proximity network. Preprint, arXiv:1211.7343.

www.risk.net/journal

- Cocco, J. F. (2009). Lending relationships in the interbank market. *Journal of Financial Intermediation* **18**(1), 24–48.
- Coleman, J. S. (1988). Social capital in the creation of human capital. American Journal of Sociology 94, S95–S120.
- Cozzo, E., Kivelä, M., De Domenico, M., Solé, A., Arenas, A., Gómez, S., Porter, M. A., and Moreno, Y. (2013). Clustering coefficients in multiplex networks. Preprint, arXiv:1307.6780.
- Craig, B., and von Peter, G. (2010). Interbank tiering and money center banks. Working Paper 332, Bank for International Settlements.
- Delpini, D., Battiston, S., Riccaboni, M., Gabbi, G., Pammolli, F., and Caldarelli, G. (2013). Evolution of controllability in interbank networks. *Scientific Reports* **3**(1626), doi:10.1038/srep01626.
- De Domenico, M., Lima, A., and Musolesi, M. (2013a). Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, **9**(6), 798–807.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S., and Arenas, A. (2013b). Mathematical formulation of multilayer networks. *Physical Review* X **3**, 041022.
- De Domenico, M., Solé-Ribalta, A., Gómez, S., and Arenas, A. (2013c). Random walks on multiplex networks. Preprint, arXiv:1306.0519.
- Elsas, R. (2005). Empirical determinants of relationship lending. *Journal of Financial Intermediation* **14**(1), 32–57.
- Elyasiani, E., and Goldberg, L. G. (2004). Relationship lending: a survey of the literature. *Journal of Economics and Business* **56**(4), 315–330 (Research Perspectives Special Issue).
- Farinha, L. A., and Santos, J. A. (2002). Switching from single to multiple bank lending relationships: determinants and implications. *Journal of Financial Intermediation* **11**(2), 124–151.
- Finger, K., Fricke, D., and Lux, T. (2013). Network analysis of the e-MID overnight money market: the international value of different aggregation levels for intrinsic dynamic processes. *Computational Management Science* **10**, 187–211.
- Fourel, V., Heam, J.-C., Salakhova, D., and Tavolaro, S. (2013). Domino effects when banks hoard liquidity: the French network. Working Paper 432, Banque de France.
- Gómez-Gardeñes, J., Reinares, I., Arenas, A., and Floría, L. M. (2012). Evolution of cooperation in multiplex networks. *Scientific Reports* 2(620), doi:10.1038/srep00620.
- Granell, C., Gómez, S., and Arenas, A. (2014). Competing spreading processes on multiplex networks: awareness and epidemics. *Physical Review* E **90**, 012808.
- Hidalgo, C. A., and Rodriguez-Sickert, C. (2008). The dynamics of a mobile phone network. Preprint, arXiv:1211.7343.
- Iori, G., De Masi, G., Precup, O. V., Gabbi, G., and Caldarelli, G. (2008). A network analysis of the Italian overnight money market. *Journal of Economic Dynamics and Control* 32(1), 259–278.
- Iori, G., Mantegna, R. N., Marotta, L., Micciche, S., Porter, J., and Tumminello, M. (2014). Networked relationships in the e-MID Interbank market: a trading model with memory. Preprint, arXiv:1403.3638.

Journal of Network Theory in Finance
- Kok, C., and Montagna, M. (2013). Multi-layered interbank model for assessing systemic risk. Technical Report, Kiel Institute for the World Economy.
- Langfield, S., Liu, Z., and Ota, T. (2014). Mapping the UK interbank system. Working Paper 516, Bank of England.
- Leicht, E., and D'Souza, R. M. (2009). Percolation on interacting networks. Preprint, arXiv:0907.0894.
- León, C., Pérez, J., and Renneboog, L. (2014). A multi-layer network of the sovereign securities market. *Borradores de Economía* **840**, 1–23.
- Little, R. G. (2002). Controlling cascading failure: understanding the vulnerabilities of interconnected infrastructures. *Journal of Urban Technology* 9(1), 109–123.
- Lozano, S., Rodriguez, X.-P., and Arenas, A. (2014). Atapuerca: evolution of scientific collaboration in an emergent large-scale research infrastructure. *Sociometrics* **98**, 1505–1520.
- Martínez-Jaramillo, S., Alexandrova-Kabadjova, B., Bravo-Benitez, B., and Solórzano-Margain, J.-P. (2014a). An empirical study of the Mexican banking system's network and its implications for systemic risk. *Journal of Economic Dynamics and Control* **40**, 242–265.
- Martínez-Jaramillo, S., Molina-Borboa, J. L., Lopez-Gallo, F., and van der Leij, M. (2014b). Trading relationships in OTC markets in Mexico. Presentation, 6th IFABS Conference on Alternative Futures for Global Banking: Competition, Regulation and Reform.
- Maslov, S., and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science* **296**(5569), 910–913.
- Menichetti, G., Remondini, D., Panzarasa, P., Mondragón, R. J., and Bianconi, G. (2014). Weighted multiplex networks. *PLoS One* **9**(6), e97857.
- Miranda, R. C., and Tabak, B. (2013). Contagion risk within firm-bank bivariate networks. Working Paper 322, Banco Central do Brasil.
- Nicosia, V., Tang, J., Mascolo, C., Musolesi, M., Russo, G., and Latora, V. (2013). Graph metrics for temporal networks. In *Temporal Networks*, Holme, P., and Saramäki, J. (eds), pp. 15–40. Understanding Complex Systems. Springer.
- Poledna, S., and Thurner, S. (2014). Elimination of systemic risk in financial networks by means of a systemic risk transaction tax. Preprint, arXiv:1401.8026.
- Poledna, S., Molina-Borboa, J. L., Martínez-Jaramillo, S., van der Leij, M., and Thurner, S. (2014). The multiplex network nature of financial systemic risk and its implications. Technical Report, Banco de México.
- Sharan, U., and Neville, J. (2007). Exploiting time-varying relationships in statistical relational models. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop* on Web Mining and Social Network Analysis, pp. 9–15. ACM.
- Solá, L., Romance, M., Criado, R., Flores, J., García del Amo, A., and Boccaletti, S. (2013). Eigenvector centrality of nodes in multiplex networks. *Chaos: An Interdisciplinary Journal* of Nonlinear Science 23(3), 033131.
- Squartini, T., and Garlaschelli, D. (2014). Stationarity, non-stationarity and early warning signals in economic networks. *Journal of Complex Networks*, doi:10.1093/comnet/ cnu012.
- Szell, M., Lambiotte, R., and Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America* **107**(31), 13636–13641.

www.risk.net/journal

Journal of Network Theory in Finance

138 J.-L. Molina-Borboa et al

- Tan, F., Xia, Y., Zhang, W., and Jin, X. (2013). Cascading failures of loads in interconnected networks under intentional attack. *Europhysics Letters* **102**(2), 28009.
- Thurner, S., and Poledna, S. (2013). DebtRank-transparency: controlling systemic risk in financial networks. *Scientific Reports* **3**(1888), doi:10.1038/srep01888.
- van Lelyveld, I., and in 't Veld, D. L. (2012). Finding the core: network structure in interbank markets. Working Paper 348, De Nederlandsche Bank.

Verbrugge, L. M. (1979). Multiplexity in adult friendships. Social Forces 57(4), 1286–1309.

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.