Robust and Sparse Banking Network Estimation

Gabriele Torri[a,b,*], Rosella Giacometti[a], Sandra Paterlini[c,d]

[a] *Department of Management, Economics and Quantitative Methods, University of Bergamo. Via dei Caniana, 2, 24127 Bergamo, Italy*
[b] *Department of Finance, VŠB-TU Ostrava. Sokolská třída 33, Ostrava, 701 21, Czech Republic*
[c] *Department of Finance and Accounting, EBS University. Gustav-Stresemann-Ring 3, 65189 Wiesbaden, Germany*
[d] *Department of Economics and Management, University of Trento. via Inama 5, I-38122 Trento, Italy*

Abstract

Network analysis is becoming a fundamental tool in the study of systemic risk and financial contagion in the banking sector. Still, the network structure must typically be estimated from noisy and aggregated data, as micro data on the status quo banking network structure are often unavailable, or the true network is unobservable. Graphical models can help researchers to infer network structures, but they are often criticized for relying too heavily on unrealistic assumptions. They also tend to yield dense structures that are difficult to interpret. Here, we propose the *tlasso* model for estimating sparse banking networks. The *tlasso* captures the conditional dependence structure between banks through partial correlations, and estimates sparse networks in which only the relevant links are identified. The model also accounts for the non-Gaussianity of financial data and it is robust to outliers and model misspecification.

Our empirical analysis focuses on estimating the dependence structure of a sample of European banks from credit default swap data. We observe that the presence of communities in the banking network plays an important role in terms of systemic risk and contagion dynamics. We also introduce a decomposition of strength centrality that allows us to better characterize the role of each bank in the network and to identify the most relevant channels for the transmission of financial distress.

Keywords: Finance, financial networks, tlasso, graphical models, CDS spreads
2010 MSC: 91G40, 91G70

*Corresponding author
Email addresses: Gabriele.torri@unibg.it (Gabriele Torri), rosella.giacometti@unibg.it (Rosella Giacometti), sandra.paterlini@ebs.edu, sandra.paterlini@unitn.it (Sandra Paterlini)

1. Introduction

The recent financial crises have pointed out the need to capture the interconnectedness between the banks in a system, in order to introduce risk-management tools capable to better control potential spillover effects in case of distress of some institutions. The "too-big-to-fail" slogan that became popular just after the 2008 financial crisis, was soon replaced by "too-interconnected-to-fail". This new expression acknowledges that systemic events can be triggered not only by the distress of large institutions, but also by small entities that cover specific roles in the system, or that are in key positions within the network. The relevance of interconnections is stressed also in the Basel Committee's regulation for assessing and identifying global systemically important banks (G-SIBs). This regulation requires that global systemic importance should be measured in terms of the impact that a bank's failure could have on the global financial system and the economy, rather than the probability that such a failure occurs. Hence, Basel introduced an indicator-based measurement approach that considers not only size but also interconnectedness, complexity, substitutability and the level of cross-jurisdictional activities (FSB, 2013).

Applications of network theory and statistical tools that can capture the dependence structure within an entire system have then flourished in finance and economics in the aftermath of the crisis, in conjunction with the rapid development of network modelling in other fields, such as sociology and biology (e.g., Albert & Barabási, 2002; Watts & Strogatz, 1998; Newman & Park, 2003). In a banking system, network nodes are typically financial institutions, and edges capture the relationship among them. The goal of these applications is to model the dynamics of financial co-movements and contagion, the potential impact on the system, and ideally to set up improved risk management tools to avoid spillover effects (see e.g., Battiston et al., 2012; Paltalidis et al., 2015; Acemoglu et al., 2015; Cont & Minca, 2016).[1]

Still, much research on the estimation of the network structure of a banking system and comparison across different modelling approaches is needed. The first contributions in the field have mostly focused on interbank lending markets, where the network structure consists in the set of bilateral credit exposures between banks (see for instance Mistrulli, 2011 and van Lelyveld & in 't Veld, 2014). Most of the initial work was set up by regulators for stress testing and to understand potential consequences of new regulation. A main limitation is still that bilateral exposures are generally non-disclosed and unavailable to researchers outside central banks. In absence of bilateral data, the network structure must be reconstructed from partial data or one must rely on other sources, such as time series of equity prices or CDS spreads.

There exists a strand of literature that reconstructs the bilateral interbank exposures using the total

[1] Other common applications of network theory in finance include the assessment of credit risk for SMEs including spatial dependence (Fernandes & Artes, 2016) and the linkages between different markets, such as Bekiros et al. (2017) that analyses the causal linkages between equity and commodity futures markets.

exposures of each bank towards the entire banking system through statistical techniques such as maximum entropy or minimum density algorithms (Elsinger et al., 2013; Anand et al., 2015). Nevertheless, such approaches usually require strong assumptions on the behaviour of the banks in the interbank market and they often fail to represent the true network structure (Mistrulli, 2011).

An alternative is instead to infer the network structure using tools that can capture co-movements and dependence patterns between financial time series to establish the existence of links among banks. An example is the estimation of credit risk networks from credit default swap (CDS) spreads or equity price times series (e.g., Puliga et al., 2014; Anufriev & Panchenko, 2015; Billio et al., 2012). One major advantage of these approaches is that they rely on public data and well known statistical modelling techniques. We focus in particular on graphical models, that are probabilistic models, nowadays widely used in machine learning (Murphy, 2012), in which the graph captures the conditional dependence structure between the nodes. Gaussian graphical models are possibly the most widely known. By relying on the assumption that nodes are random variables with a multivariate normal distribution, edges are estimated as partial correlations in these models. Values of zero imply then independence between nodes, while the conditional probability among nodes factorizes according to the partial correlation graph (Lauritzen, 1996). The advantage of using conditional dependence, compared to direct correlations, is that it allows us to single out the direct co-movements between nodes, controlling for spurious connections due to common exposures. Our paper aims to fill some gaps in the literature by proposing a robust method, based on graphical models and regularization techniques, for estimating only the relevant links in the network from CDS spread time series data. The estimated network then provides information on the dependence among banks and possible paths of shock propagation. We underline that we consider synchronous, and not delayed co-movements of time series, and that we estimate sparse undirected networks. As a consequence, we cannot interpret the edges as causal relationships, at least in a Granger (i.e. intertemporal) sense. Other approaches, such as *Granger causality networks* (Billio et al., 2012), variance decomposition (Diebold & Yılmaz, 2014) and transfer entropy (Bekiros et al., 2017) would allow us to consider intertemporal and directed relationships by modeling the data, either in a Vector AutoRegression (VAR) framework or by using an informational theoretical approach. Still, we believe that given the aim of our work, partial correlation networks deliver an appropriate estimate of the connectivity patterns among banks since, assuming that the markets are sufficiently liquid and promptly reacting, the information available to the market should be discounted by the prices fast enough to be reflected by synchronous relationships.[2] Although it is not always possible to match the co-movement networks to the actual

[2]Intertemporal causation may indeed indicate the presence of financial turmoil, as reported by Billio et al. (2012), that show how Granger causality networks may allow to identify periods of market dislocation and distress, and Jenkins et al. (2016) who report that during the financial crisis in 2008, the US CDS market was less efficient (in terms of semi-strong market efficiency).

network structure of the interbank market, some studies highlight their usefulness to capture salient aspects of the financial system, which allows regulators to design more effective policies and to set-up risk monitoring and mitigating tools (Anufriev & Panchenko, 2015). Other studies found some interdependence between co-movement networks and interbank exposures networks. It is the case of Abbassi et al. (2017), that highlight the capability of detecting similar patterns when comparing partial correlation networks based on CDS data in the German banking system and the actual bilateral exposures provided by the Deutsche Bundesbank credit register, suggesting that market based measures can serve well in the absence of bilateral interbank market data.

One of the main drawbacks of Gaussian graphical models, besides the normality assumption, which is hardly satisfied by financial times series, is the fact that the estimated graph is typically dense and difficult to interpret. Regularization methods can then be exploited to detect sparse graph structures for which not all the edges are active and only the relevant links are detected. The *glasso* model, one of the best known approaches for matrix regularization in the statistical literature, induces sparsity in the network structure by introducing a penalty proportional to the 1-norm of the precision matrix in the estimation process (Friedman et al., 2008). While *glasso* has found a range of applications in biology and other fields, its use within the financial literature is still limited (e.g., Goto & Xu, 2015; Anufriev & Panchenko, 2015). This is possibly because the assumption of Gaussianity is often too restrictive when dealing with financial times series.

Here, we introduce the *tlasso*, a model that allows to estimate the partial correlation matrix by relying on the more realistic assumption of multivariate t-Student variables. This allows us to better capture the leptokurtic behaviour of financial times series. The model is also more robust to model misspecification and the presence of outliers compared to *glasso*, providing more stable and reliable estimates of the sparse network structure. However, differently from the Gaussian setting, under the t-Student distributional assumption the absence of an edges does not imply conditional independence, but only conditional zero correlation. To the best of our knowledge, besides Torri et al. (2017) on portfolio selection and optimization, *tlasso* has never been applied to the estimation of the networks of financial institutions. After illustrating the robustness properties of *tlasso* compared to *glasso* in a simulation study, we focus on estimating the network structure of a sample of large European banks, providing some financial interpretation of the network in two ways. First, we analyse the structural properties of the system using a relevant set of network measures. Second, we propose a decomposition of the *strength centrality* measure to characterize more accurately the role of each bank in the network and to highlight the most relevant channels for the transmission of financial distress. We thus introduce a tool that can be used to evaluate the risk profile of each bank in the system.

Our empirical results suggest a highly connected network structure, characterized by geographical communities and by the absence of a core-periphery structure. The analysis complements previous

studies that focus on domestic banking systems (e.g., Anand et al., 2015; Mistrulli, 2011; van Lelyveld & in 't Veld, 2014), which are generally characterized by very sparse, strongly tiered network structures with a small set of large banks that play the role of hubs. Other works conducted on samples of international financial institutions, while relying on different modelling techniques, provide results that are in line with ours, although they put less or no emphasis on the community structure. In particular, Craig & Saldías (2016) found that, considering a network derived from equity prices, the European banking system is characterized by a high level of interconnectivity, high clustering coefficient, and high assortativity. Our results are also in line with Aldasoro & Alves (2016), which studied the interbank network of large European financial institutions, and found evidence of a high network density and a core composed by a large number of banks.

The paper is organized as follows: Section 2 describes the *tlasso* model, after introducing the Gaussian graphical model and *glasso*. Section 3 illustrates the properties of *tlasso* and compares it to *glasso* in a simulation set-up. Section 4 reports the empirical results obtained by analysing a sample of large European banks using CDS data. Estimation results are presented along with network properties, followed by the proposal of the decomposition of *strength centrality* and a discussion of the financial and economic implications of our analysis. Section 5 draws the main conclusions.

2. The *tlasso* model

2.1. Preliminaries

We briefly review partial correlation networks in the context of Gaussian graphical models and the *glasso* model that, starting from a dense network, identifies a *sparse* partial correlation matrix under the hypothesis of Gaussianity.

Let $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(m)})$ be a random vector with a multivariate Gaussian distribution $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix. We can define an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes in $\mathcal{V}$ correspond to each element of $\boldsymbol{X}$, the edges $\mathcal{E}$ consist of the pairs of random variables with non-zero partial correlations, $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$, and the edge weights correspond to partial correlations ρ_{ij}. The partial correlations are computed from the inverse of the covariance matrix $\boldsymbol{\Omega} := \boldsymbol{\Sigma}^{-1}$ (i.e., the precision matrix) as follows:

$$\rho_{ij} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \qquad i,j = 1, \ldots, m,$$

where $\{\omega_{ij}\}$ is an element of the matrix $\boldsymbol{\Omega}$. The estimation of the graph structure corresponds to the estimation of the precision matrix $\boldsymbol{\Omega}$. The partial correlation matrix $\boldsymbol{P}$ is:

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \boldsymbol{\Omega} \boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}}, \tag{1}$$

where $\boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} = diag(\frac{1}{\sqrt{\omega_{ii}}})$ and $\boldsymbol{I}$ is a conformable identity matrix. The representation of partial correlations in the matrix form $\boldsymbol{P}$ is the *weighted adjacency matrix* of the graph, that is, the $m \times m$ matrix

in which the entries are the weights of the edges in the network. Notice that, following a common convention, the diagonal elements of $\boldsymbol{P}$ are equal to zero. The reader is referred to Lauritzen (1996) for Gaussian graphical models, while Anufriev & Panchenko (2015) provide a review of partial correlation networks in the Gaussian framework.

Traditional estimation techniques for graphical models, such as maximum likelihood, typically return dense adjacency matrices in which all the elements off the main diagonal are different from zero. It may then be challenging to identify the relevant links among nodes. Moreover, the presence of collinearity in the data would result in ill-conditioned estimates of the covariance matrices and, consequently, large estimation error for the precision matrix (Torri et al., 2017).

The *glasso* model, introduced by Friedman et al. (2008), was proposed as a tool to solve this issue and to obtain *sparse* estimates of the precision matrix $\boldsymbol{\Omega}$. Under the assumption of Gaussianity, the precision matrix can be estimated by penalized quasi maximum likelihood, where the penalty is proportional to the 1-norm of the precision matrix. Omitting constants and multiplicative factors, the estimates obtained with *glasso* can be computed by solving the following optimization problem:

$$\hat{\boldsymbol{\Omega}}_{glasso} = \arg\max_{\boldsymbol{\Omega}}(log|\boldsymbol{\Omega}| - tr(\boldsymbol{\Omega}\boldsymbol{\Sigma}_S) - \lambda||\boldsymbol{\Omega}||_1), \tag{2}$$

where $\boldsymbol{\Sigma}_S$ is the empirical sample covariance matrix, λ is the regularization parameter (the larger the λ, the sparser the estimated model), $|\cdot|$ is the determinant of a matrix, $tr(\cdot)$ the trace and $||\cdot||_1$ the 1-norm. The optimization problem can be solved efficiently using the algorithm proposed by Friedman et al. (2008).

2.2. The tlasso Model

One of the main limitations of *glasso* is the Gaussianity assumption that, as widely discussed in the literature, generally does not hold for financial time series, typically characterized by fat tails and leptokurtic distributions (e.g., Cont, 2001). We then consider *tlasso*, a recently introduced model for the inference of sparse partial correlation networks under the assumption that the data follow a multivariate t-Student distribution (Finegold & Drton, 2011). Compared to *glasso*, the distributional assumptions of *tlasso* are less restrictive for financial applications. Furthermore, the use of t-Student distributional assumption as an alternative to Gaussianity is known to increase the robustness of the estimates in presence of outliers in several applications, such as linear and nonlinear regression, robust estimation of the mean and covariance matrix with missing data, unbalanced multivariate repeated-measures and multivariate nonlinear regression Lange et al. (1989).

Let $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(m)})$ be an m-variate random vector with a multivariate t-Student distribution $t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v)$, with v degrees of freedom, mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Psi}^{-1}$ ($m \times m$ positive

semi-definite matrix[3]). The covariance and precision matrices are then $\boldsymbol{\Sigma} = \frac{v}{v-2}\boldsymbol{\Psi}^{-1}$ and $\boldsymbol{\Omega} = \frac{v-2}{v}\boldsymbol{\Psi}$, respectively. Similarly to the Gaussian case, we can associate to the distribution a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, in which $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$, where $\{\rho_{ij}\}$ are the elements of the partial correlation matrix $\boldsymbol{P}$ computed from the precision matrix $\boldsymbol{\Omega}$ as in (1). The *tlasso* model allows us to estimate a sparse $\widehat{\boldsymbol{\Psi}}_{tlasso}$ and the corresponding partial correlation matrix $\widehat{\boldsymbol{P}}$ that represents the network structure.

In particular, we consider a scale-mixture representation of the t-Student distribution. Let $\boldsymbol{W} \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Psi}^{-1})$ and $\tau \sim \Gamma(v/2, v/2)$ be random variables distributed as a multivariate Gaussian and a gamma distribution, respectively. Then:

$$\boldsymbol{X} = \boldsymbol{\mu} + \frac{\boldsymbol{W}}{\sqrt{\tau}} \sim t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v), \tag{3}$$

where $\boldsymbol{\Psi}^{-1}$ is the dispersion matrix of the t-Student distribution, $\boldsymbol{\mu}$ is a $m \times 1$ vector, and v is the number of degrees of freedom in the multivariate t-Student distribution (Kotz & Nadarajah, 2004).[4]

In contrast to the Gaussian set-up, when using the t-Student distribution, or the class of elliptical distributions in general, an absence of correlation does not necessarily correspond to conditional independence (Baba et al., 2004). However, despite the lack of conditional independence for $\omega_{ij} = 0$, it can be proved that if two nodes j and k are separated by a set of nodes C in $\mathcal{G}$, then $X^{(j)}$ and $X^{(k)}$ are conditionally uncorrelated given $X^{(C)}$ (Finegold & Drton, 2011). It is then reasonable to substitute the conditional independence with zero partial correlation or zero conditional correlation. In this case, disconnected vertices in a graphical model can be considered orthogonal to each other after the effects of other variables are removed. The absence of conditional correlation entails that a mean-square error optimal prediction of variable $X^{(j)}$ can be based on the variables $X^{(k)}$, which correspond to neighbours of the node j in the graph.

2.2.1. The Expectation-Maximization Algorithm

The *tlasso* model can be estimated using an Expectation-Maximization (EM) algorithm (Finegold & Drton, 2011). The EM estimation procedure exploits the scale-mixture representation (3), and it is computationally efficient since it is based on the *glasso* algorithm, which is iteratively applied at every M-step of the algorithm.

The EM algorithm treats τ as a hidden variable in the E-step, exploiting the fact that the conditional distribution of $\boldsymbol{X}$ given τ is $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Psi}^{-1}/\tau)$. Then, in the M-step the algorithm maximizes the penalized log-likelihood of the latent Gaussian vector using the *glasso* procedure.

[3]We choose the uncommon parametrization based on the inverse of the diffusion matrix to highlight the relevance of Ψ in the graphical model.

[4]Note that the scale-mixture representation clarifies how the use of t-Student distribution leads to more robust inference, as extreme observations can arise from small values of τ.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be an n-sample drawn from $t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v)$. The EM algorithm iterates the following two steps:

- **E-step**
  - Given:

$$\mathbb{E}[\tau|\boldsymbol{X} = \mathbf{x}] = \frac{v + m}{v + (\delta_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Psi}))}, \tag{4}$$

    where $\delta_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Psi} (\mathbf{x} - \boldsymbol{\mu})$, from the current estimates $\widehat{\boldsymbol{\mu}}^{(t)}$ and $\widehat{\boldsymbol{\Psi}}^{(t)}$, we compute $\widehat{\tau}^{(t+1)}$ for the $(t+1)$th iteration:

$$\widehat{\tau}_i^{(t+1)} = \frac{v + m}{v + (\delta_{\boldsymbol{X}_i}(\widehat{\boldsymbol{\mu}}^{(t)}, \widehat{\boldsymbol{\Psi}}^{(t)}))} \qquad i = 1, \ldots, n \tag{5}$$

- **M-step**
  - Compute the estimates at iteration $t + 1$:

$$\widehat{\boldsymbol{\mu}}^{(t+1)} = \frac{\sum_{i=1}^n \widehat{\tau}_i^{(t+1)} \boldsymbol{X}_i}{\sum_{i=1}^n \widehat{\tau}_i^{(t+1)}}, \tag{6}$$

$$\widehat{\boldsymbol{S}}^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \widehat{\tau}_i^{(t+1)} [\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}}^{(t+1)}][\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}}^{(t+1)}]' \tag{7}$$

  - Then, the estimate $\widehat{\boldsymbol{\Psi}}^{(t+1)}$ is computed by solving the following optimization problem:

$$\widehat{\boldsymbol{\Psi}}^{(t+1)} = \arg\max_{\boldsymbol{\Psi}} \left(log|\boldsymbol{\Psi}| - tr(\boldsymbol{\Psi} \widehat{\boldsymbol{S}}^{(t+1)}) - \lambda ||\boldsymbol{\Psi}||_1 \right), \tag{8}$$

    which relates to (2) for *glasso* and can be solved using the algorithm from Friedman et al. (2008).

The E and M steps are sequentially iterated until a convergence criterion is satisfied, that is, until the maximum term in absolute value of the matrix difference between $\widehat{\boldsymbol{\Psi}}$s in two consecutive iterations is smaller than a given threshold. While convergence to a stationary point is guaranteed in the penalized versions of EM (McLachlan & Krishnan, 2007), the algorithm is not guaranteed to find the global maximum since the penalized log-likelihood function to be maximized is not concave (Finegold & Drton, 2011).

2.3. The Regularization Parameter

The regularization parameter λ controls the intensity of the shrinkage towards zero of individual elements of the precision matrix: the larger λ, the more sparse is the precision matrix estimate. Different approaches, such as cross-validation and information criteria, can be used to select the optimal regularization parameter λ (Finegold & Drton, 2011). We rely on the Bayesian Information Criterion (BIC) for

the calibration of λ, as the literature highlights that it performs better than cross-validation (Foygel & Drton, 2010).

For *tlasso*, we consider a grid of values of λs, that is $\lambda_i \in \mathcal{C} = [\lambda_{min}, \ldots, \lambda_{max}]$ and estimate the *tlasso* network for each λ value. Then, we choose the value of λ that solves the BIC optimization problem:

$$\min_{\lambda \in \mathcal{C}} BIC = -2log(\hat{L}_{\lambda_i}) + k \times log(n), \tag{9}$$

where $\hat{L}_{\lambda_i}$ is the value of the likelihood function for the multivariate t-Student distribution with the parameters computed using the corresponding λ_i parameter, k is the number of active edges in the network, and n is the number of observations. The calibration procedure is analogous for *glasso*, with the only difference being the use of the likelihood function for a multivariate Gaussian distribution instead of a multivariate t-Student distribution. In preliminary analyses on simulated data, we also compared the BIC with the Extended Bayesian Information Criterion (EBIC) and the Akaike Information Criterion (AIC), finding that the BIC provides more accurate results than the others.

3. Simulation Analysis

As a first step, we test the *tlasso* on simulated data using several network configurations and distributions, and we compare its performances to the ones of the *glasso*. We consider 16 settings with different combinations of network structures and underlying distributions of the data (see Section 3.1). For each combination, we perform 30 Monte Carlo runs and estimate the networks using *tlasso* and *glasso* on a grid of 30 different λ values ranging from $\lambda_{min} = 0.007$ to $\lambda_{max} = 1$. We set the number of degrees of freedom to $v = 3$, the lowest possible value that yields a finite covariance matrix, to test for the influence of fat tails. We also investigate how *tlasso* and *glasso* estimates are affected by the sample size, comparing estimates obtained for $n = 100$, 200, 500 and 1000. Finally, we evaluate the performance of the BIC for the optimal choice of the regularization parameter λ.

3.1. Simulation Set-up

In the simulation set-up, we consider four distributional assumptions and four network configurations, chosen from commonly used benchmark cases and more complex structures that are typical of financial networks. We then run the simulation analysis for the sixteen combinations of network structures and distributions. We describe the structure of the network in terms of the precision matrix $\boldsymbol{\Omega}$, which makes it possible to easily derive the partial correlation matrix $\boldsymbol{P}$ as in (1).

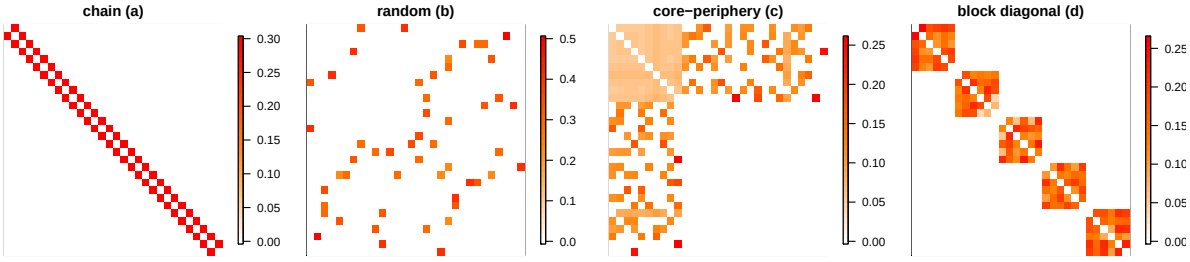
a. **Chain graph**: This simple graph structure can be represented by a matrix $\boldsymbol{\Omega}_a$, where all the non-diagonal entries are zero except the elements on the first sub- and super-diagonal. In particular, these entries have been set equal to -0.3, while diagonal elements are set equal to 1.

b. **Sparse random graph**: The matrix $\mathbf{\Omega}_b$ is defined as follows: first a value of -1 is assigned independently with probability 0.05 to the elements in the lower diagonal, then the matrix is mirrored to obtain a symmetric adjacency matrix. A value equal to the negative of the sum of the elements in the respective columns minus one is assigned to the elements on the diagonal. This guarantees that the matrix is positive definite.

c. **Core-periphery matrix:** The matrix $\mathbf{\Omega}_c$ is consistent with a core-periphery structure, that is, it has a group of strongly interconnected nodes (the core) and a periphery connected to nodes within the core, but not among nodes within the periphery. The core is composed of 10 elements and the percentage of the active edges between the core and the periphery has been set to 30%.

d. **Block diagonal matrix**: The matrix $\mathbf{\Omega}_d$ is characterized by a strong block diagonal structure with five distinct blocks. This structure is consistent with a network composed of five disconnected sub-networks. In particular, the blocks are modelled independently and in each of them the time series are generated as a linear combination of two independent factors plus some random noise.

Figure 1: Alternative network structures based on partial correlation matrices



The figure represents the color coded heatmaps of the adjacency matrices $\boldsymbol{P}$ corresponding to network structure (a - d). The structures are in order: the chain graph (a), a random graph with high sparsity (b), a core-periphery graph (c) and a block diagonal graph (d).

We assume the following four underlying distributions:

1. **Multivariate Gaussian distribution**: $X \sim \mathcal{N}_m(0, \mathbf{\Omega}^{-1})$
2. **Multivariate t-Student distribution**: $X \sim t_m(0, \frac{v-2}{v}\mathbf{\Omega}^{-1}, v), \qquad \text{with } v = 3$
3. **Contaminated Gaussian distribution**: Normal distribution with covariance matrix $\mathbf{\Omega}^{-1}$ where 15% of the observations are outliers, drawn from independent normal distributions with larger variances:

$$X \sim \mathcal{N}_m(0, \mathbf{\Omega}^{-1}) \times b + \mathcal{N}_m(0, 2 \times diag(\mathbf{\Omega}^{-1})) \times (1 - b),$$

$$\text{with } b \sim Bernoulli(p), \text{and} \quad p = 0.85.$$

4. **Empirical marginals from real CDS data and *t-copula***: Distribution characterized by a *multivariate t-copula* and marginals estimated from real CDS data using a Gaussian kernel. The data are described in Section 4 and refer to the period 01/01/2009 – 30/06/2016.

3.2. Performance Measures

We first evaluate the performances of *tlasso* and *glasso* estimators in terms of ROC (Receiver Operating Characteristics) curves, measuring their ability to identifying correctly the edges in the network. A good estimator should have a high sensitivity (high ratio of true positives to all positives), while having a high specificity (high ratio of true negatives to all negatives), that is, it should identify all the active edges in the original network without introducing spurious ones. ROC curves are a powerful and commonly used tool for the identification of these features. For binary classifiers, they consist in a plot of the sensitivity rate over (1 - specificity) for different thresholds (see for instance Hanley & McNeil, 1982). We use them for the performance analysis of *glasso* and *tlasso*, comparing network estimates with different levels of sparsity as determined for a grid of λ values. In order to provide better representation of the dispersion of the data, instead of computing the average value over several runs and drawing an interpolated ROC curve, we represent all the individual data points.

In addition to ROC curves, we evaluate the quality of the estimates of the partial correlation matrix $\boldsymbol{P}$ using the Frobenius norm as distance measure between the optimal estimate of the adjacency matrix $\widehat{\boldsymbol{P}}$ and the known adjacency matrix $\boldsymbol{P}$, as follows:

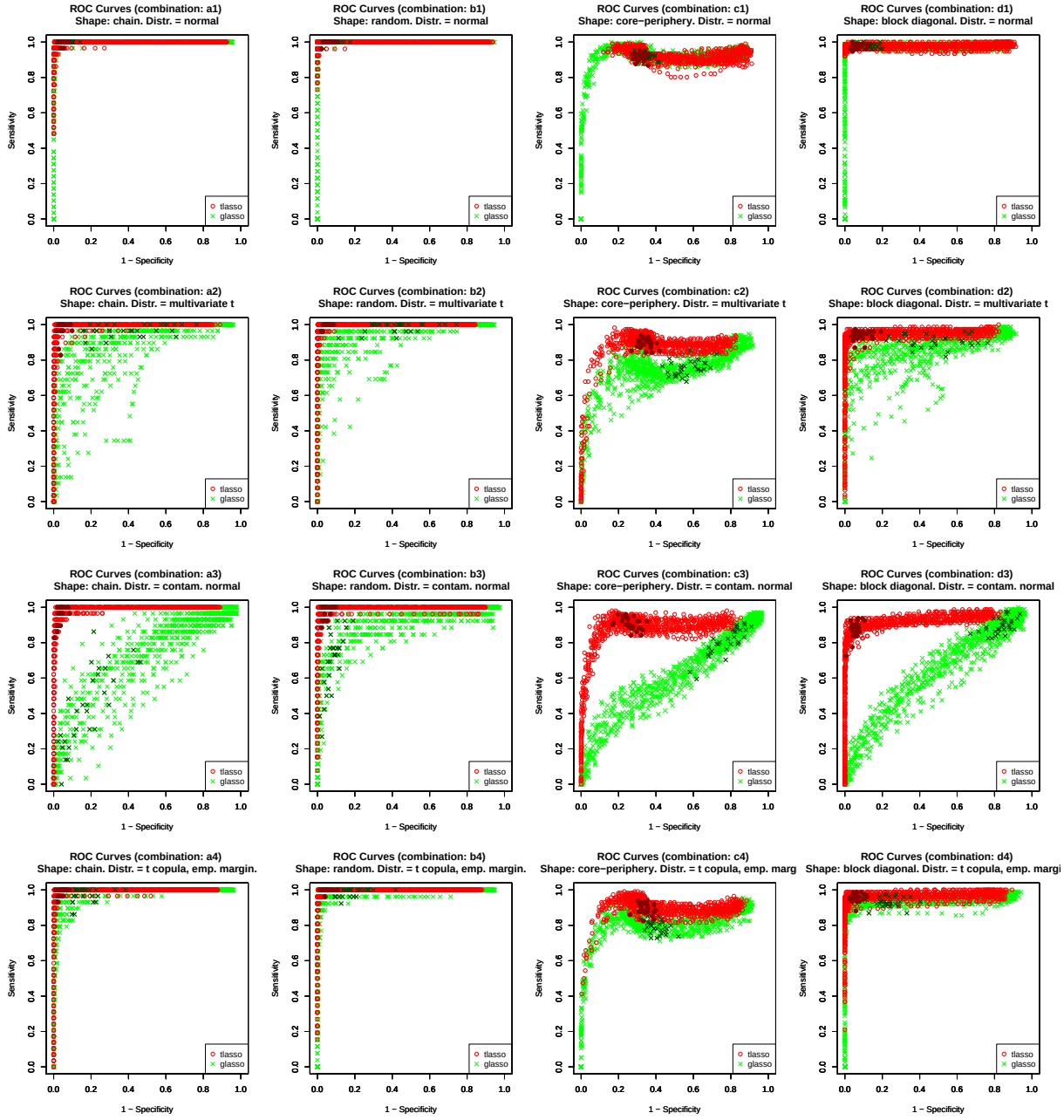
$$D_F(\widehat{\boldsymbol{P}}, \boldsymbol{P}) = ||\widehat{\boldsymbol{P}} - \boldsymbol{P}||_F = \sqrt{(tr((\widehat{\boldsymbol{P}} - \boldsymbol{P})(\widehat{\boldsymbol{P}} - \boldsymbol{P})'))}, \tag{10}$$

where $\widehat{\boldsymbol{P}}$ is the estimate of the partial correlation matrix and $\boldsymbol{P}$ is the one used for the generation of the data. In particular, we use the Frobenius norm to measure the accuracy of the estimation of the optimal networks identified by BIC.

3.3. Simulation Results

Figure 2 displays the ROC curves for all the four combinations of network configurations (from a to d) and distributions of the data (from 1 to 4), described in Section 3.1. Each data point refers to one estimate of the network structure for a different value of λ. Red circles refer to *tlasso* networks, while green crosses indicate *glasso* ones. The optimal networks identified by the BIC are highlighted with darker colours. The more concentrated the data are in the upper left corner, the better the performance, as measured by a high number number of true positives and a low number of false positives. We see that the quality of the estimates varies drastically across different combinations of distributions and network structures. Moving from top to bottom, we analyse the performance of *glasso* and *tlasso* on data with different distributions. In the case of data with a Gaussian distribution (first row), the performances

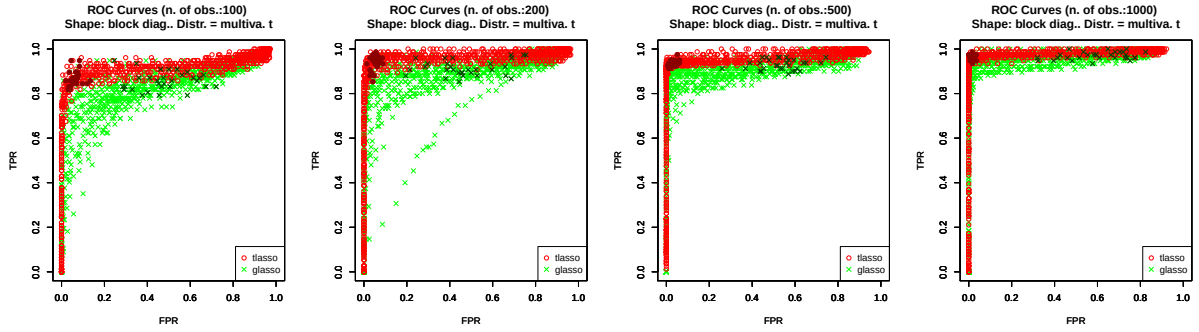
Figure 2: ROC curves for 16 simulation set-ups



ROC curves for *tlasso* and *glasso* estimation with different network structures and underlying distributions, as defined in Section 3.1 (structures (a - d), from left to right; distributions (1 - 4), from top to bottom). Red circles refer to *tlasso* and green crosses to *glasso*. Dark green circles and dark red crosses refer to the optimal estimates identified using the BIC for *tlasso* and *glasso*, respectively.

of *glasso* and *tlasso* are similar, though the first model is correctly specified while the second is not. When the distribution is not Gaussian (rows 2, 3 and 4), *tlasso* performs consistently better than *glasso*,

Figure 3: ROC curves for different sample size



ROC curves for *tlasso* and *glasso* estimation with block diagonal network structure (d) and multivariate t-Student distribution with 3 degrees of freedom (2). Green points refers to *glasso* and red points to *tlasso*. Darker points are the optimal estimates identified using the BIC. From left to right the four graphs display results for sample size equal to 100, 200, 500 and 1000.

especially in the case of the distribution characterized by the presence of outliers (row 3).

When studying the effect of the partial correlation structure (Figure 2 from left to right), we notice that the topology of the network strongly affects the quality of the estimates provided by the *tlasso* and *glasso*. We see that both models perform better in case of simple structures like the chain graph (a) or a very sparse random graph (b). They have problems, however, in correctly identifying more complex structures such as the core-periphery (c) or the block diagonal configurations (d).

The optimal networks identified by the BIC are highlighted with darker circles and crosses. We see that, in the case of the *tlasso*, the BIC is highly effective for all the specifications considered, as it often selects optimal λs corresponding to good estimates of the true network structure (high sensitivity and specificity, lying in the upper-left corner of the graph). In the case of the *glasso*, the BIC works well for the selection of the optimal parameter λ when the model is correctly specified (row 1), but performs poorly when the distribution is not Gaussian or in presence of outliers (rows 2, 3 and 4).

Figure 3 illustrates as a robustness check how the quality of the estimates changes with the sample size $n = 100, 200, 500$ and 1000. From all the combinations, we report only the case of a known block diagonal network structure and multivariate t-Student distribution with 3 degrees of freedom (corresponding to combination (d2) in Figure 2). Other pairs provide similar results, which are available upon request. As expected, the larger the sample size n, the lower the estimation error, as indicated by the concentration of the points in the upper left corner in Figure 3 for $n = 1000$. We see that *tlasso* consistently outperforms *glasso*, which fails to provide accurate estimates, even with large sample size, due to the model misspecification and its lack of robustness.

Finally, Table 1 reports the average Frobenius distance of the optimal networks from the true networks for *tlasso* (Panel a) and *glasso* (Panel b). The results confirm those of the ROC curve analysis and show that *tlasso* outperforms *glasso* in most of the specifications, obtaining smaller distances. The difference

between the models is particularly relevant for the distribution 3, which is characterized by outliers. The only exceptions are the simulations with normally distributed data, where *glasso* and *tlasso* show similar performance.

In summary, the simulation study shows that *tlasso* generally outperforms *glasso*, in particular under non-normality, model misspecification, and in presence of outliers.

Table 1: Average Frobenius distance for *tlasso*

	Network structure			
Distribution	Chain (a)	Random (b)	Block diag. (c)	Core-periphery (d)
Panel a - *tlasso*				
Multivariate normal (1)	1.22	1.05	1.11	0.71
(s.d.)	(0.10)	(0.12)	(0.06)	(0.04)
Multivariate t (2)	1.29	1.09	1.13	0.79
(s.d.)	(0.22)	(0.08)	(0.06)	(0.08)
Contam. normal (3)	1.51	1.32	1.16	0.87
(s.d.)	(0.15)	(0.12)	(0.06)	(0.06)
t-copula + emp. marg. (4)	1.24	1.11	1.17	0.76
(s.d.)	(0.13)	(0.11)	(0.06)	(0.05)
Panel b - *glasso*				
Multivariate normal (1)	1.15	0.95	1.12	0.73
(s.d.)	(0.11)	(0.09)	(0.05)	(0.05)
Multivariate t (2)	1.88	1.67	1.91	1.72
(s.d.)	(0.31)	(0.28)	(0.30)	(0.40)
Contam. normal (3)	2.21	1.95	3.29	3.45
(s.d.)	(0.09)	(0.15)	(0.35)	(0.37)
t-copula + emp. marg. (4)	1.38	1.28	1.53	1.15
(s.d.)	(0.10)	(0.11)	(0.11)	(0.13)

Average Frobenius distance (and standard deviation in parenthesis) between optimal estimators $\widehat{\boldsymbol{P}}$ and known network structure $\boldsymbol{P}$ for *tlasso* (Panel a) and *glasso* (Panel b) for all network structures and distributions presented in Section 3.1. Optimal networks are selected using BIC.

4. Empirical Analysis

Here, we present the application of the *tlasso* model to the problem of estimating the network structure of a sample of large European banks. Our goal is to infer the network structure from the partial

correlations between CDS to capture dependence related to credit risk, and to describe the properties of the system and its evolution over time. We further provide insights on the network estimates by the *tlasso* when compared to those obtained using the *glasso*. Finally, we introduce a decomposition of *strength centrality* that allows us to characterize accurately the role of each bank in the network and to highlight the most relevant channels for the transmission of financial distress.

4.1. Banking Data

The dataset consists of 31 weekly time series of CDS spreads (5 years maturity, quoted in Euros) of European banks from 12 countries. The data have been downloaded from Thomson Reuters Datastream and span the time period from 01/01/2009 to 30/06/2016. 20 out of 31 banks belong to countries within the Eurozone, while the other 11 are located in the United Kingdom, Sweden and Denmark. Our database includes 85% of those banks with total assets over EUR 500 billion under the ECB's supervision in 2016 and 47% of the banks involved in the European Banking Authority (EBA) stress-test exercise of 2016. The complete list of the banks is reported in Appendix A. For the analysis, we consider the log-differences of the CDS spreads.

The *glasso* and *tlasso* models assume that the data follow a multivariate Gaussian and t-Student distribution, respectively. Due to complexity in testing the multivariate goodness of fit in high dimensionality (see e.g., Justel et al., 1997; McAssey, 2013), we do not test directly for the multivariate distributional assumption, but instead we focus on the marginal univariate and bivariate distributions.[5] Results for individual banks are reported in Appendix A. On average, the marginal distributions show low skewness and high excess kurtosis. We test for two parametric distributions: the t-Student and the Gaussian, fitting them using maximum likelihood. We then evaluate the goodness of fit using the Kolmogorov-Smirnov test and the χ^2 goodness of fit test, both at a significance level of $\alpha = 1\%$. According to the Kolmogorov-Smirnov test, we do not reject the null hypothesis of a t-Student distribution for any of the banks except for one (i.e., Bayerische Landesbank), while the assumption of Gaussianity is rejected for 8 of the 31 banks. The χ^2 goodness of fit test rejects the null hypothesis of a t-Student and Gaussian distribution for 4 and 18 banks, respectively. Concerning the fitting of t-Student marginal distributions, the estimated numbers of degrees of freedom are low, ranging from 1.75 to 8.80 with mean value equal to 5.37. The statistical tests thus provide evidence in favour of t-Student marginal distributions.

Concerning the dependence structure, we analyse 465 bivariate distributions for each pair of banks, fitting different copula families and selecting the best-fitted copula model for each pair using the Bayesian

[5] Following Kotz & Nadarajah (2004), any partition of a random variable that follows a multivariate t-Student distribution is also distributed as a t-Student. By testing the distribution of marginals we find a necessary, although not sufficient, condition for the entire distribution to follow a t-Student distribution. Such property holds also for multivariate Gaussian distributions as a special case.

Information Criterion (BIC). The copulas we consider are *Gaussian*, *t* (with 3, 6 and 10 degrees of freedom), *Gumbel*, and *reverse Gumbel*. This selection encompasses models with different levels of tail dependence and allows us to model both symmetric and asymmetric relationships. Table 2 reports the percentages of the bivariate distributions for which each copula is the best fit according to the BIC. The vast majority of the tests provide evidence in favour of the *t-copula*. In fact, the *t-copula* with 6 degrees of freedom turns out to be optimal for more than half of the pairs (54.4%), and *t-copulas* with 3, 6 or 10 degrees of freedom are optimal in 93.2% of the cases. The bivariate data therefore appear to display a relatively high level of tail dependence.

Table 2: Bivariate copula fitting for pairs of banks.

Copula	Percentage
t-copula (3 degrees of freedom)	10.8%
t-copula (6 degrees of freedom)	54.4%
t-copula (10 degrees of freedom)	28.0%
Gaussian copula	4.7%
Gumbel copula	1.9%
Reverse Gumbel copula	0.2%

Percentage out of the 465 bivariate relationships between each pair of banks' CDS spreads, for which each copula is the best fit according to BIC (01/01/2009 – 30/06/2016).

The results obtained for the marginal and bivariate distributions suggest the need to consider a methodology that moves beyond the Gaussian distributional and dependence assumption. We therefore expect the *tlasso* model to be a more appropriate tool than the *glasso* due to empirical properties of the data and the *tlasso*'s robustness to outliers and model misspecification.

The analysis of the autocorrelograms of the log-differences of CDS spreads and their squared returns (not reported for brevity) does not highlight relevant evidence of serial correlation and heteroskedasticity. In the rest of the analysis we therefore consider the variables as independent and identically distributed (i.i.d.). As a robustness check, in Appendix C we repeat the analysis on the residuals of an ARMA-GARCH model using a static conditional correlation model (CCC-GARCH) and a dynamic conditional correlation one (DCC-GARCH) (see Appendix C). The results obtained with these specifications are very similar to the baseline scenario.

We finally study the model after controlling for conditioning variables, considering as factors a set of 11 European sovereign CDS spreads (i.e., Austria, Belgium, Denmark, France, Germany, Italy, Netherland, Portugal, Spain, Sweden and United Kingdom), and re-estimating the model after controlling for their effect. Also in this case, the results are consistent to ones obtained with raw data. The results are reported in Appendix C.

4.2. Stability and Robustness to Outliers

As illustrated in Section 3, *tlasso* outperforms *glasso* in the majority of the simulation set-ups we consider. Here, we test the performance of *glasso* and *tlasso* on real data, using them to estimate the sparse correlation network underlying the European banking system.

Since the true network structure is unknown, we cannot use the performance measures described in Section 3 for the simulated data. Instead, we compare the *glasso* and *tlasso* focusing on the stability and robustness to outliers. In particular, we consider the stability of the network estimates in a rolling window analysis, computing measures of similarity between estimated networks on different time-windows. The rationale behind this approach is that a robust method should return very similar network estimates on largely overlapping time-windows and the presence of sudden changes in the network structure would indicate that the model is not robust to outliers and prone to estimation errors.

We consider the period from 01/01/2008 to 30/06/2016, with rolling windows of 100 weekly observations for a total of 343 windows.[6] In order to allow for a more meaningful comparison between time periods, we do not calibrate the penalization parameters λ for *tlasso* and *glasso* in each window, but instead we set the parameter in such a way that the average density of the network across the time windows is equal for the two models. For the estimation of *tlasso* we set the degrees of freedom to 5, close to the mean value of the marginal distributions of the data (see Appendix A).

We use the Frobenius norm of the difference between the two adjacency matrices, defined as in (10) as an indicator for the distance between networks. Figure 4 represents graphically the distance between each pair of rolling windows and table 3 reports the average distance computed on windows 5, 10 and 15 weeks apart.

The results show that the Frobenius distance between networks tends to increases with the distance in time of the estimation windows, which is in line with our expectations. This can be seen in Figure 4, where the points close to the diagonal (which represent windows close to each other) have smaller Frobenius distance. Similarly, we see in Table 3 that networks estimated on windows 15 weeks apart from each other differ more than those estimated on windows 10 and 5 weeks apart. Comparing the two models, we notice that, on average, the distance between networks estimated on different windows is smaller for *tlasso* than for *glasso*, resulting in a more stable structure over time for *tlasso*. Interestingly, the *tlasso* networks are not only more stable on average, but they are also characterized by a smoother transition of the network over the rolling windows. This can be observed in Figure 4, where *tlasso* has a less block-diagonal structure than *glasso*. Such result is confirmed in Table 3, where the value of the standard deviation of the Frobenius distances for *tlasso* is much smaller than for *glasso*. The absence of sudden changes suggests that *tlasso* is less affected by outliers and extreme events, which leads to more

[6]For the rolling analysis the sample is reduced to 29 banks, since CDS series for *Intesa San Paolo* and *Unicredit* are not available for the entire period.

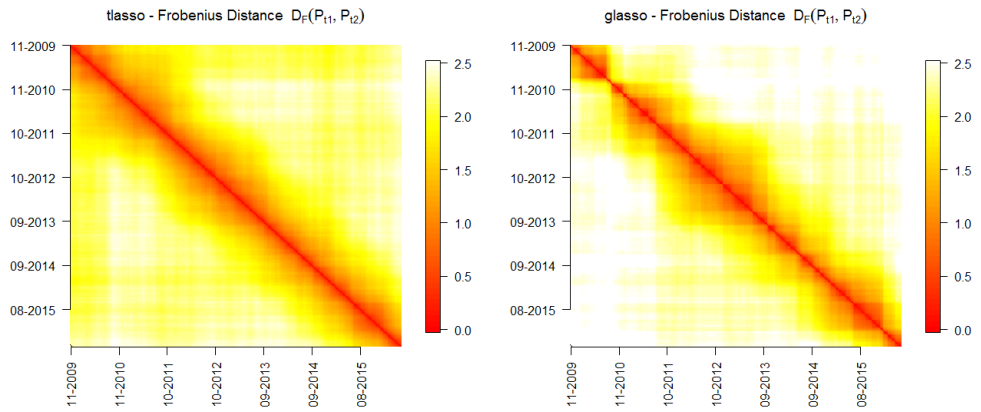
reliable and stable estimates.

Table 3: Frobenius distance on rolling window networks

	Frobenius distance	
	tlasso	*glasso*
5 weeks	0.55	0.64
	(0.06)	(0.23)
10 weeks	0.78	0.91
	(0.07)	(0.26)
15 weeks	0.95	1.11
	(0.09)	(0.27)

Average Frobenius distance and standard deviation (in brakets) between networks computed on rolling windows respectively 5, 10 and 15 weeks apart.

Figure 4: Frobenius distance between networks on different time windows.



Frobenius distance between networks computed on different time windows. Lower distance denotes more similar networks. The horizontal and vertical axes represent the last day in each time window.

4.3. Structural Analysis of the European Banking System Network

As a next step, we focus on the identification of the structural properties of the European banking system by computing relevant network indicators and monitoring their changes in two sub-periods: *during-crisis* (01/01/2009 – 31/12/2012) and *post-crisis* (01/01/2013 – 30/06/2016). We focus on the meso-scale properties of the network, such as the presence of a core-periphery or a community structure (see Borgatti & Everett, 2000; Fortunato, 2010).

Most of the available literature analyse the banking systems in individual countries, for which the bilateral exposures in the interbank markets are more easily available (e.g., Craig & Von Peter, 2014

for Germany, van Lelyveld & in 't Veld, 2014 for the Netherlands, and Mistrulli, 2011 for Italy). The network structures in these studies are typically characterized by a high level of sparsity, a core-periphery structure, a scale-free configuration (i.e., networks where degree and strength distributions are power-law), the presence of hubs (i.e., nodes with a number of edges that greatly exceed the average) and *disassortative mixing* patterns (i.e., highly connected banks tend to connect to peripheral ones). The properties of international financial systems are less well-studied, and seem to be characterized by more complex structures. Aldasoro & Alves (2016) analyse the interbank market bilateral exposures of 53 large European banks and identify a highly interconnected and layered structure with a core-periphery configuration, characterized by a large core. Craig & Saldías (2016) use equity data to analyse the global banking system and find a complex hierarchical structure, with a relevant regional homophily, a rich-club phenomenon (highly connected nodes tend to be mutually linked) and a core-periphery structure. Here, we provide further evidence in estimating and interpreting dependence networks based on CDS data in international banking systems.

We first look at the descriptive network statistics, reported in Table 4. We then compute a set of indicators useful to describe the meso-scale properties of the network (see core-periphery structure, presence of geographical communities and assortative mixing). A detailed description of the network indicators is reported in Appendix B.

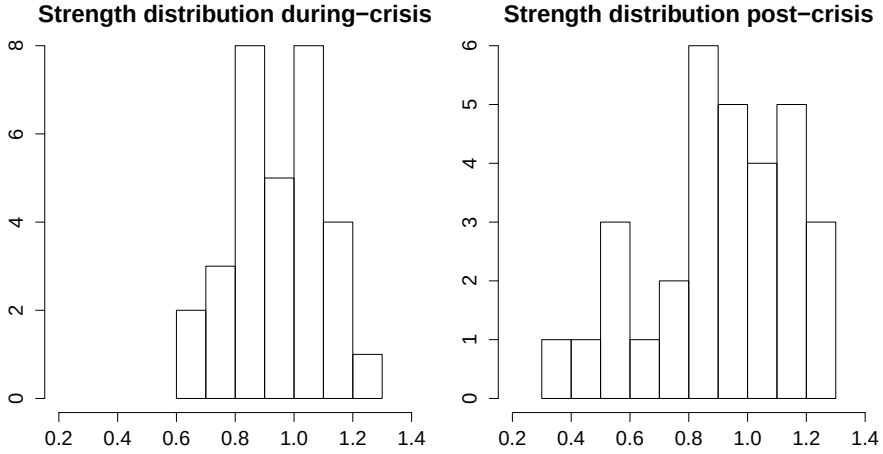
Table 4: Descriptive network statistics for CDS Networks

	during-crisis	post-crisis
Density	49.25%	52.26%
Average strength	0.95	0.91
Largest connected component	100%	100%
Positive edges %	87.77%	79.42%
Positive edges % (weighted)	96.59%	89.28%

Descriptive network indicators computed on *tlasso networks* for the *during-crisis* and the *post-crisis* periods. See Appendix B for the definition of network indicators.

In Table 4 we notice that for both sub-periods the network is connected (it exists at least a path that connect any two nodes in the network) and by similar values of *density* (49.25% and 52.26%), *average degree* (14.77 and 15.68), and *average strength* (0.95% and 0.91%), suggesting a highly interconnected structure. Edges with positive weights are the majority, especially in the *during-crisis* period and when computed in weighted terms. Figure 5 reports the *strength distribution* for the sub-periods. Concerning the *during-crisis* period (left plot), we notice that the support is narrower than in the *post-crisis* period (right plot) and that the distribution is approximately symmetric, suggesting that the networks do not have a scale-free structure and are not characterized by the presence of hubs. In the *post-crisis*

Figure 5: Strength distribution of the estimated networks



Strength distribution in the *during-crisis* (left) and *post-crisis* (right) periods.

period, the *strength distribution* is negatively skewed, indicating the presence of some banks that are less interconnected to the system compared to the *during-crisis* period.

Table 5: Network indicators for *tlasso*

	during-crisis		post-crisis	
	tlasso	random network	*tlasso*	random network
Freeman Centralization	0.47**	0.66	0.52*	0.69
Coreness	0.11***	0.34	0.15***	0.31
Modularity (wrt country)	0.36***	-0.04	0.29***	-0.03
Clustering coefficient	2.70%**	2.12%	2.94%***	2.39%
Assortativity (wrt eigen. centr.)	0.61***	0.35	0.35	0.34
Assortativity (wrt str. centr.)	-0.05	0.08	0.06	0.09
Assortativity (wrt betw. centr.)	-0.05	-0.07	-0.08	-0.07
Assortativity (wrt Bonacich centr.)	0.03	0.08	0.11	0.08

Network indicators computed on *tlasso* networks for the *during-crisis* and the *post-crisis* periods. The Table also reports the median value of the indicators computed on 1000 random rewirings of the networks. The stars denote the statistical significance based on the empirical distribution of the indicators on the random networks (***, **, * denote p-value smaller than 1%, 5% and 10%, respectively).

Table 5 reports a set of network indicators chosen to highlight the meso-scale structural properties of the networks, focusing in particular on the presence of core-periphery and community structure. As a reference, Table 5 also shows the median value of the indicator for a random network and the statistical

significance computed on the quantiles of the distribution of the indicators on a random network.[7]

The *Freeman centralization* provides information about the extent to which the network is centred around its most central node. A high centralization is typically associated with a tiered structure or with the presence of strong hubs. In the *during-crisis* period, this indicator is statistically significantly lower than in the random networks in both sub-periods, suggesting the absence of a *core-periphery* structure. This is confirmed by the low *coreness*,[8] equal to 0.11 in the *during-crisis* period and to 0.15 in the *post-crisis* period.

The following two indicators, *modularity* and the *clustering coefficient*, allow us to measure the tendency of the networks to be organized in communities. In particular, the *modularity* computed on the country partition describes the tendency of the networks to form communities aligned with national borders. The indicator has a high and statistically significant value, especially in the *during-crisis* periods, where it takes a value of 0.36. The presence of communities is also suggested by the high value of the *clustering coefficient*,[9] that is typically associated to networks with communities (Fagiolo, 2007).

Finally, we compute the *assortativity*, an indicator that measures the tendency of nodes to connect to similar ones. In particular, we consider the *assortativity* with respect to four different centrality measures (*eigenvector-*, *strength-*, *betweenness-* and *Bonacich power centrality*) to check the tendency of nodes to be linked to nodes with the same level of importance. Negative assortativity is typical of network with hubs, and such systems are typically resilient to random failures, but fragile to specific attack to hubs Watts & Strogatz (1998). Positive assortativity, instead, is typical of social networks. Overall, the network does not present neither an assortative or disassortative behaviour. The only exception is the *eigenvector centrality* in the *during-crisis* period, where the indicator is statistically significantly higher than in a random network. The absence of disassortative mixing may seem in contrast with the literature, that shows disassortativity in financial networks (see e.g., Mistrulli, 2011; Hurd, 2016). This may be explained considering that we included in our analysis only the largest European banks, discarding the smallest, and arguably less interconnected institutions. The result may also be related to the presence of a community structure. In fact, Newman & Park (2003) shows that communities can generate a higher than expected *assortative mixing* since nodes in larger and denser groups have a higher number of connections relative

[7] The random networks have been computed by a *rewiring* procedure that consists in randomly selecting two edges (a,b) and (c,d) and substituting them with (a,d), (b,c). The rewiring procedure is repeated iteratively until the network is completely randomized (see for instance Fortunato, 2010). This procedure allows us to destroy existing structural properties while maintaining the *degree* of the nodes and other features such as the *average strength*. The significance level is obtained from the quantiles of the empirical distribution of the indicators computed on 1000 randomly rewired networks.

[8] *Coreness* is an indicator that denotes the tendency of a network to have a strongly connected core and a sparsely connected periphery (Borgatti & Everett, 2000).

[9] The *clustering coefficient* is an indicator that denotes the tendency of the network to "create triangles", that is, when a node a is connected to nodes b and c, there is a high chance that b and c are connected.

to nodes than in smaller communities. This would lead to highly interconnected nodes being connected to other highly connected nodes, increasing the *assortativity*.

Figure 6 shows a visual representation of the networks using a force layout color-coded by country. The representation confirms the structural properties suggested by the indicators: the banking system is characterized by a high interconnection and a relevant community structure aligned with national division, especially in the *during-crisis* period.

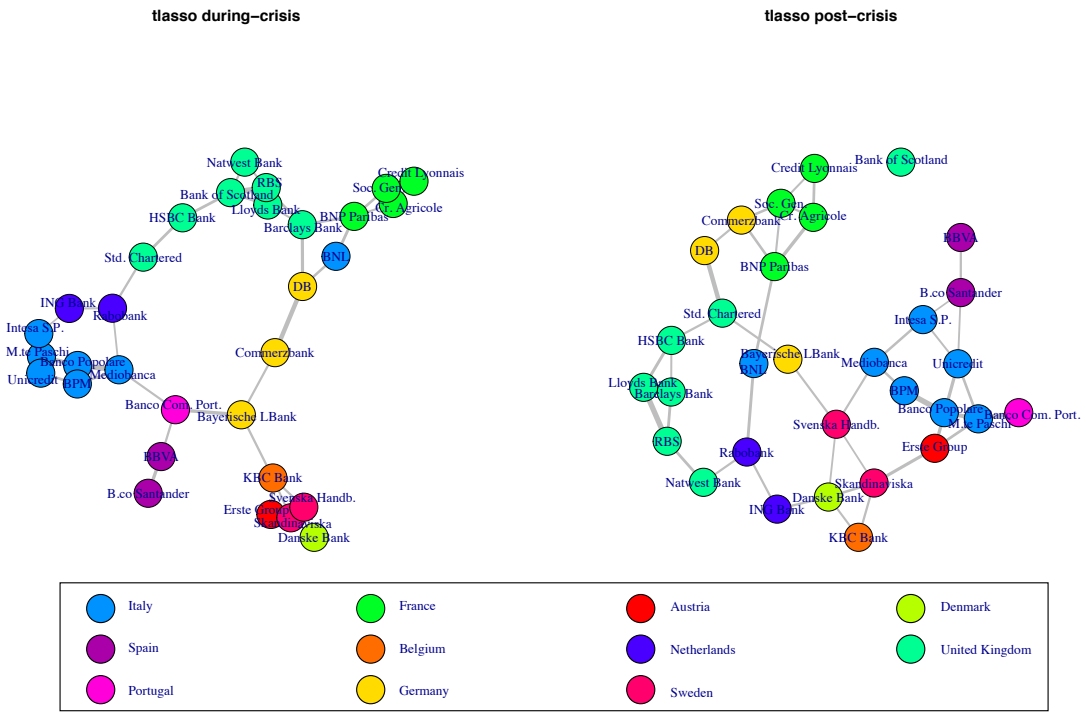
The comparison with existing literature is complicated by the differences in methodology and sample selection. As we mentioned before, the structure identified here differs significantly from that identified in studies that focus on actual bilateral exposures in national banking systems. They describe much sparser networks, scale-free distributions for the degree and strength, core-periphery configurations and disassortative mixing (Mistrulli, 2011). Such divergence can be explained by the fact that in our analysis we include only large banks that are part of the *core* of the network in each individual country. The comparison with studies focused on international banking networks shows more consistent results. Aldasoro & Alves (2016), using the algorithm from Anand et al. (2015) on a dataset of bilateral exposure between large European banks, find a large core composed by twenty to thirty banks, depending on the type of interconnections considered and Craig & Von Peter (2014) underline the presence of a strongly interconnected structure in Europe.

To our knowledge, our work is the first to put emphasis on the presence of geographical communities in the European banking system. Aldasoro & Alves (2016) did not provide indicators to measure this feature and Craig & Von Peter (2014) only describe qualitatively the presence of regional homophily and the higher density of the network structure within countries. From an economic point of view, the identification of national communities is consistent with the different national legal and economic systems, and possibly related to the European sovereign debt crisis and CDS pricing.

It is well known in the complex network literature that, in general, structural and topological properties influence the diffusion of processes on a network and strongly affect the vulnerability to failures and attacks (Watts & Strogatz, 1998; Newman, 2002; Fortunato, 2010). Concerning banking systems, the structural properties have clear implications in terms of diffusion of financial contagion and the assessment of the network properties is useful for both financial institutions and policy makers. The literature agrees that the presence of a highly interconnected structure in the interbank market increases the stability of the system, but that in presence of liquidity shocks, it makes the system less stable, leading to a *robust-yet-fragile* property. Moreover, the relationship between structural properties and stability is complex and non linear (Hurd, 2016; Chinazzi & Fagiolo, 2015). Acemoglu et al. (2015), for instance, shows that more densely connected financial networks enhance financial stability but, beyond a certain point, dense interconnections lead to a less stable financial system. Hurd (2016) shows that the presence of *assortative mixing* facilitates the diffusion of contagion in the system, and Hübsch & Walther (2017) show how the

presence of inhomogeneities among banks can enhance contagion risk. The modelization of the diffusion of financial contagion in networks partitioned in communities is still an open issue in the literature. Applications in other fields show that the presence of communities influences the diffusion of processes over a network (see Weng et al., 2013; Galstyan & Cohen, 2007). We thus expect that this feature may be relevant also in financial applications. Finally, we underline that the presence of communities, similarly to other structural network properties, may have an influence on the interpretation of centrality indicators and raises new questions on how to measure the importance of a node in a system. In the next section, we try to address these questions by introducing a decomposition of *strength centrality*.

Figure 6: *tlasso* networks - graphical representation



Graphical representations of the *tlasso* networks using a force layout. For clarity, only the edges corresponding to a partial correlation larger than 0.1 in absolute value are depicted.

4.4. Centrality Measures for Networks With Communities

In the previous section, we analysed the structural properties of the estimated networks of the European banking system using the *tlasso* model. Here, we shift the focus to the analysis of individual institutions, studying their interconnectedness and their systemic relevance. Following the definition reported by Schwarcz (2008), systemic risk is closely related to the propagation and reverberation of risk in a system, and in the literature it has been typically measured either by econometric indicators or

by a network approach. Concerning the econometric approach, several studies developed indicators that measure risk trasmission and spillover to the system and from the system, such as *marginal expected shortfall* (Engle & Brownlees, 2010), *expected capital shortfall* (Acharya et al., 2012), $CoVaR$ and $\Delta CoVaR$ (Adrian & Brunnermeier, 2016). Concerning to network approaches, they typically aim at assessing the relevance of an institution in a system using traditional centrality measures such as *eigenvector-*, *strength-* and *betweeness centrality* (see for instance Alter et al., 2015), or by introducing specific measures such as *debtrank* by Battiston et al. (2012).

A common feature of most of the measures in the literature, either econometric or network-based, is that they summarize in a single measurement the relationship between a given institution and the system, but often do not allow to characterize the relationship between an individual bank's systemic risk and the topological features of the entire network. We introduce a methodology that allows to take explicitly into consideration the network topology, in particular the presence of geographical communities while assessing the role of each individual institution. To clarify the concept, we can think of a network with a geographical community structure. In such network, a bank may be characterized by high centrality because it has strong international interconnections or because it is strongly connected to banks in the same country. We can imagine that these two situations can be very different for regulators who want to measure the systemic importance of a bank or understand the level of EU integration. Our proposed method provides a flexible tool for dealing with this situation by introducing a decomposition of *strength centrality*. This measure, when computed on partial correlation networks, has an interesting interpretation in terms of shock diffusion, that we discuss below.

We first recall the interpretation of the partial correlation matrix in terms of *hedge regressions*, that are a set of regressions in which each asset is regressed against all others (see Stevens, 1998). Assuming that $\mathbb{E}[\boldsymbol{X}] = \mathbf{0}$ (where $\mathbf{0}$ is a conformable vector of zeros), the hedge regressions are expressed as:

$$X_t^{(i)} = \sum_{j \neq i} \beta_j^{(i)} X_t^{(i)} + \varepsilon_t^{(i)} \qquad i = 1, \ldots, m, \tag{11}$$

where $X^{(i)}$ is the ith asset, $\beta_j^{(i)}$ the coefficient for the jth asset in the ith regression and $\varepsilon^{(i)}$ the residuals. Or in matrix form

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{\varepsilon}, \tag{12}$$

where $\boldsymbol{B}$ is a matrix with off-diagonal elements ij equal to $\beta_j^{(i)}$ and diagonal elements equal to 0.

We can relate the elements ω_{ij} of $\boldsymbol{\Omega}$, the precision matrix of the data, to the βs as

$$\omega_{ij} = \begin{cases} -\dfrac{\beta_j^{(i)}}{\nu_i} & \text{if } i \neq j \\ \dfrac{1}{\nu_i} & \text{if } i = j \end{cases}, \tag{13}$$

where ν_i is the variance of $\varepsilon^{(i)}$. Using (12), (13), and (1), after some algebra we have

$$\boldsymbol{X} = \boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \boldsymbol{P} \boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}} \boldsymbol{X} + \boldsymbol{\varepsilon}. \tag{14}$$

In order to account for the different conditional variances in the $X^{(i)}$s, we introduce the rescaled variable $\mathbf{x} = \boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}}\boldsymbol{X}$, in such a way that the rescaled residuals $\mathbf{e} = \boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}}\boldsymbol{\varepsilon}$ have unit variance. By premultiplying the terms in (14) by $\boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}}$ we have

$$\mathbf{x} = \boldsymbol{P}\mathbf{x} + \mathbf{e}. \tag{15}$$

The rescaling allows to remove the effects of variables endogenous to the network by expressing the extent of the shock in terms of the conditional variances (see Anufriev & Panchenko, 2015).

Considering the regression interpretation presented here, we now use the network structure estimated by *tlasso* to study the systemic risk and financial contagion. As we mentioned before, the setting described so far does not allow to identify direct causation between variables, since it considers only synchronous co-movements. Similarly to Anufriev & Panchenko (2015), we then consider an observational interpretation of (15), where the residuals $\mathbf{e}$ can be seen as an exogenous shock, and the partial correlation network edges as the channels of propagation: a shock $\mathbf{e}$ hitting a set of nodes will have an effect on the other nodes in virtue of (15) and the effect on the direct neighbours of the hit nodes is computed as $\boldsymbol{P}\mathbf{e}$ (*first round effect*).

We can then compute the effect on the entire system in terms of a unitary shock hitting the ith node in terms of *first round effect* as

$$c_i^S = \sum_{i \neq j} \alpha_j \rho_{ij} = \boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i \tag{16}$$

where $\mathbf{e}_i$ is a unitary shock hitting node i and $\boldsymbol{\alpha}$ is a vector denoting the weight of each node in the system. Notice that in case of $\boldsymbol{\alpha} = \mathbf{1}$, where $\mathbf{1}$ is a conformable vector of ones, the quantity c_i^S consists in the *strength centrality* of node i. For the symmetry of $\boldsymbol{P}$, the network is not directed and the quantities can also be interpreted as the effect of a unitary shock hitting all the nodes on the node i.

As (16) is an homogeneous functions of degree 1 with respect to $\boldsymbol{\alpha}$, we can then use the following Euler decomposition:

$$c_i^S = \boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i = \sum_{j=1}^{m} \boldsymbol{\alpha}_j \frac{\partial \boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i}{\partial \boldsymbol{\alpha}_j} = \boldsymbol{\alpha}' \nabla (\boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i), \tag{17}$$

where ∇ denotes the gradient with respect to vector $\boldsymbol{\alpha}$. We emphasize that this framework accounts only for the immediate transmission of shock. Once the shock is transmitted to the banks connected to the first one, it could be propagated further, generating a cascade effect. We could account for these propagation effects by considering different *rounds* of transmission of the shock, similarly to Anufriev & Panchenko (2015) or Battiston et al. (2012). Here, we consider only the first *round* of contagion transmission, and focus on the short term effects. We leave the modeling of successive steps for further research, also considering that the structural properties of the network may change after a shock hits due

to reactions of banks and interventions from banking authorities.[10]

These components can then be aggregated according to meaningful attributes of the data, in our case the geographical location of the banks.

We apply the decomposition of *strength centrality* to the networks estimated using *tlasso*. In order to make the results easier to read, and to allow for a better comparison between banks, we perform the strength decomposition according to three broad geographical groups and not on individual countries. The regions we consider are *Southern European* countries (i.e., Italy, Spain and Portugal), *Central European* countries (i.e., Germany, France, Austria, Belgium and Netherlands) and the countries *outside the Eurozone* (i.e., United Kingdom, Sweden and Denmark).[11]

Table 6 shows the average decomposition of the strength in each geographical area. On average, the largest part of the *strength centrality* of a bank can be attributed to connections with banks in the same geographical region. This confirms the presence of a community structure aligned to geographical divisions, especially in *Southern Europe* and for the group of countries outside the Eurozone, for which the edges within the same group in the crisis-period account respectively for 69.19% and the 68.83% of the total *strength centrality*. Instead, the banks in *Central European countries* seem to be characterized on average by a slightly less regional interconnectedness, as for them edges directed to banks in the same area account on average for 57.98% of the *strength centrality*. The result would be even more pronounced if we exclude the French banks, that are tighly connected to each other. The decomposition yields similar results for the two periods, denoting stability in the structure of European banking network. Still, in the *post-crisis* period the banks show on average a higher international connectivity, quantified by a lower share of *strength centrality* attributable to banks in the same geographical group, especially for the banks outside the Eurozone.

Figure 7 provides the relative decomposition of *strength centrality* for individual banks (top panels), the value of *strength centality* (middle panels) and *eigenvector centrality* (bottom panels). The decomposition allows us to highlight the specific position of each bank in the system, and to describe more accurately its properties. We see for instance that Italian and French banks are characterized by a localized interconnectivity, while German, Danish and Dutch banks have on average stronger connections to banks in other geographical areas. The representation allows also to easily identify outliers and banks with specific features, such as the Italian bank *BNL*, whose connections are mostly towards *Central European* banks. This is not surprising, as BNL is part of the French *BNP Paribas* group. The identification

[10]The effect of the shock after all the reverberations can be measured as the *Bonacich power centrality* (Anufriev & Panchenko, 2015), and it can be decomposed as in (17). However, due to the reverberations, the shocks diffuse more uniformly to the entire networks, making the decomposition less informative. The results are not reported for brevity, and are available upon request.

[11]The decomposition of *strength centrality* for individual countries gives results consistent with those reported below. They are available from the authors upon request.

of the most internationalized banks can be used to detect potential *bridges* for the diffusion of contagion between distant areas. Such banks may then be monitored more accurately by regulators. We also show how the decomposition of the *strength centrality* can be used to highlight specificities of each node related to the structural properties of the network. For instance, focusing on *RBS* (Royal Bank of Scotland), we notice that in the *during-crisis* period, it is one of the institutes with the highest *eigenvector centrality*, but its interconnectivity structure spans mostly a limited geographical area. According to the estimated network structure, a potential shock hitting RBS would therefore spread initially to a localized neighbourhood, while a shock on a bank with a wider interconnectivity structure may potentially affect international institutions faster.

Table 6: Decomposition of *strength centrality* - average % of strength directed to each geographical area

	During-crisis			Post-crisis		
	Southern	Central	Outside	Southern	Central	Outside
Southern	69.19 %	23.42 %	7.84 %	67.20 %	21.17 %	12.06 %
Central	21.05 %	57.98 %	21.82 %	20.19 %	56.52 %	27.85 %
Outside	7.60 %	22.21 %	68.83 %	10.85 %	23.44 %	60.75 %

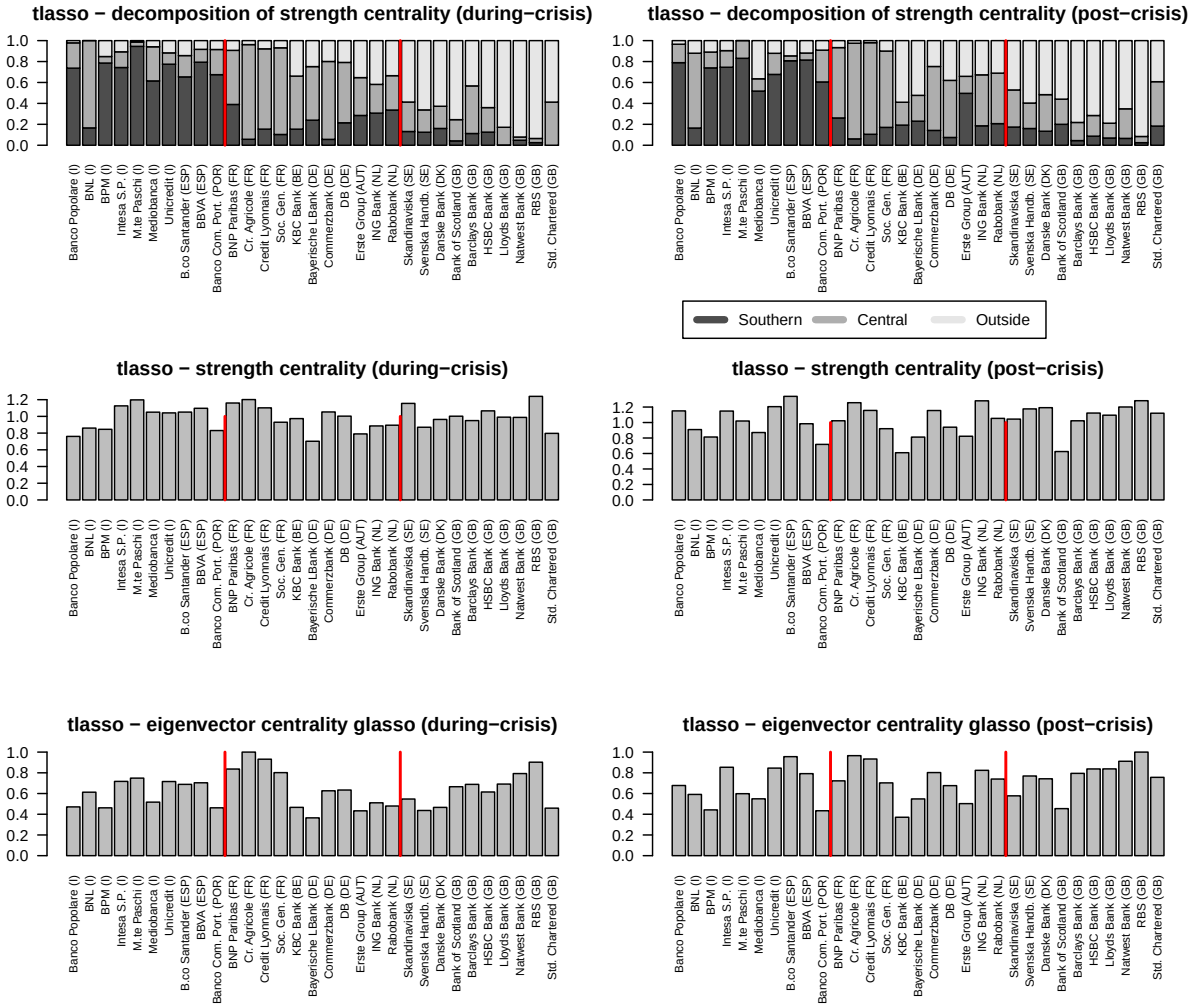
Decomposition of strength centrality. The Table reports the average value of the strength directed to each geographical area. *Southern*: Southern European countries (i.e., Italy, Spain and Portugal). *Central*: Central European countries (i.e., Germany, France, Austria, Belgium and Netherlands). *Outside*: Countries outside the Eurozone (i.e., United Kingdom, Sweden and Denmark).

We point out that the interaction between the structural properties of the network and the local centrality measures assumes a particular relevance for the development of macroprudential regulation and supervision toolkits, that aim to protect the financial system as a whole, in contrast to microprudential regulation, which focuses on the stability of individual institutions, as discussed by Freixas et al. (2015).

In this context, the *strength centrality* decomposition that we propose can be used to complement other network based centrality measures in order to define a new multi-dimensional approach for the identification of systemic relevance, in a spirit similar to the G-SIB framework, an indicator-based assessment methodology introduced in November 2011 by the Financial Stability Board as a response to the 2008 global financial crisis (FSB, 2013).

The question whether a bank with more localized interconnections poses a smaller or greater threat to the system in terms of contagion risk is still an open issue in the literature and should be further studied. On the one hand, a localized system may slow down the diffusion of distress, allowing the contagion to dissipate before it reaches all the nodes in the network. On the other hand, the concentration of distress in a limited group of banks may reduce the system's ability to dissipate shocks and may trigger feedback effects that would amplify the initial shock. Still, the *strength centrality* decomposition that we propose can provide a useful tool for the characterization of the contagion channels for each bank. The study of

Figure 7: Decomposition of *strength centrality* and centrality measures.



Geographical decomposition of *strength centrality* (upper panels) and value of *strength centrality* (middle panels) and *eigenvector centrality* (lower panels). The three geographical areas (*Southern Europe*, *Central Europe* and *outside Eurozone*) are highlighted with red vertical bands. *Southern*: Southern European countries (i.e., Italy, Spain and Portugal). *Central*: Central European countries (i.e., Germany, France, Austria, Belgium and Netherlands). *Outside*: Countries outside the Eurozone (i.e., United Kingdom, Sweden and Denmark).

contagion mechanisms in networks with geographical communities is high on our agenda.

5. Conclusion

In this work, we use partial correlation networks to estimate the structure of the European banking system from CDS data, introducing the *tlasso* model in the financial network literature. Such model estimates efficiently sparse partial correlation networks under the assumption of a multivariate t-Student distribution, inducing sparsity by means of a 1-norm matrix penalization. We show by a simulation

analysis and on real CDS data that the *tlasso* is a suitable tool for the estimation of the European banking system network, as it performs better than the alternative *glasso* model when the data are non Gaussian and it is robust to model misspecification and to the presence of outliers.

We then study the topological properties of the financial networks estimated using *tlasso*, focusing on the meso-scale properties. We find evidence of a highly interconnected network characterized by a relevant community structure based on geographical divisions, especially during the crisis period. The network is also characterized by the absence of a core-periphery structure and limited evidence of *assortative mixing*. The results contribute to the growing literature that analyses the network topology of the European banking systems and is consistent with previous works (Craig & Saldías, 2016; Aldasoro & Alves, 2016). Instead, as expected, there are relevant differences compared to studies that focus on national banking systems, where banking networks are typically characterized by the presence of hubs and core-periphery structures.

Finally, we propose a novel decomposition of the *strength centrality* measure, that can be used to assess the contagion risk exposure of financial institutions and to account for the topological features of the network. We use the decomposition to analyse the centrality of financial institutions in presence of geographical communities, and we observe that banks from *Southern Europe* and from *Outside the Euro-zone* on average have a more geographically localized interconnectivity structure, while *Central European* banks have on average a more internationalized set of interconnections. The presence of geographical communities emerges both during and after the crisis, and appears to be more prominent in the crisis period.

Our study is, to our knowledge, one of the first to put such emphasis on the presence of a geographical community structure, to quantify the geographical centrality of each bank in an international banking system. Further research will focus on the analysis of the mechanisms of contagion diffusion in a network with communities, and the implications in terms of macroprudential supervision policies.

Acknowledgement

Sandra Paterlini acknowledges ICT COST Action IC1408 from CRoNoS. Gabriele Torri acknowledges the support of the Czech Science Foundation (GACR) under project 15-23699S and SP2017/32, an SGS research project of VSB-TU Ostrava. Rosella Giacometti and Gabriele Torri acknowledge the support given by University of Bergamo research funds 2016 2017.

References

Abbassi, P., Brownlees, C., Hans, C., & Podlich, N. (2017). Credit risk interconnectedness: What does the market really know? *Journal of Financial Stability*, *29*, 1–12.

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *The American Economic Review*, *105*, 564–608.

Acharya, V., Engle, R., & Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *The American Economic Review*, *102*, 59–64.

Adrian, T., & Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, *106*, 1705–1741.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.*, *74*, 47–97.

Aldasoro, I., & Alves, I. (2016). Multiplex interbank networks and systemic importance: An application to European data. *Journal of Financial Stability*, *in press*. doi:10.1016/j.jfs.2016.12.008.

Alter, A., Craig, B. R., & Raupach, P. (2015). Centrality-based capital allocations. *International Journal of Central Banking*, *11*, 329–377.

Anand, K., Craig, B., & Von Peter, G. (2015). Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance*, *15*, 625–636.

Anufriev, M., & Panchenko, V. (2015). Connecting the dots: Econometric methods for uncovering networks with an application to the Australian financial institutions. *Journal of Banking & Finance*, *61*, 241–255.

Baba, K., Shibata, R., & Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, *46*, 657–664.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 3747–3752.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., & Caldarelli, G. (2012). Debtrank: Too central to fail? Financial networks, the FED and systemic risk. *Scientific Reports*, *2*, 541.

Bekiros, S., Nguyen, D. K., Junior, L. S., & Uddin, G. S. (2017). Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *European Journal of Operational Research*, *256*, 945–961.

Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, *104*, 535–559.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics*, *72*, 498–505.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*, 1170–1182.

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, *21*, 375–395.

Chinazzi, M., & Fagiolo, G. (2015). Systemic risk, contagion, and financial networks: A survey. Available at SSRN 2243504.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, *1*, 223–236.

Cont, R., & Minca, A. (2016). Credit default swaps and systemic risk. *Annals of Operations Research*, *247*, 523–547.

Craig, B., & Saldías, M. (2016). *Spatial Dependence and Data-Driven Networks of International Banks*. International Monetary Fund, WP/16/184.

Craig, B., & Von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, *23*, 322–347.

Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*, 119–134.

Elsinger, H., Lehar, A., Summer, M., Fouque, J., & Langsam, J. (2013). Network models and systemic risk assessment. *Handbook on Systemic Risk*, *1*, 287–305.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*, 339–350.

Engle, R., & Brownlees, C. (2010). Volatility, correlation and tails for systemic risk measurement. Manuscript, Stern School of Business, New York University.

Fagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E*, *76*, 026107.

Fernandes, G. B., & Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*, *249*, 517–524.

Finegold, M., & Drton, M. (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *The Annals of Applied Statistics*, *5*, 1057–1080.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*, 75–174.

Foygel, R., & Drton, M. (2010). Extended bayesian information criteria for gaussian graphical models. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*, 215–239.

Freixas, X., Laeven, L., & Peydró, J.-L. (2015). *Systemic risk, crises, and macroprudential regulation*. MIT Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441.

FSB (2013). *Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement*. Technical Report. URL: `http://www.bis.org/publ/bcbs255.pdf`.

Galstyan, A., & Cohen, P. (2007). Cascading dynamics in modular networks. *Physical Review E*, *75*, 036109.

Gómez, S., Jensen, P., & Arenas, A. (2009). Analysis of community structure in networks of correlated data. *Physical Review E*, *80*, 016114.

Goto, S., & Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, *50*, 1415–1441.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36.

Hübsch, A., & Walther, U. (2017). The impact of network inhomogeneities on contagion and system stability. *Annals of Operations Research*, *254*, 61–87.

Hurd, T. R. (2016). *Contagion! Systemic Risk in Financial Networks*. Springer.

Jenkins, N. T., Kimbrough, M. D., & Wang, J. (2016). The extent of informational efficiency in the credit default swap market: evidence from post-earnings announcement returns. *Review of Quantitative Finance and Accounting*, *46*, 725–761.

Justel, A., Peña, D., & Zamar, R. (1997). A multivariate kolmogorov-smirnov test of goodness of fit. *Statistics & Probability Letters*, *35*, 251–259.

Kotz, S., & Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.

Lange, K. L., Little, R. J., & Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, *84*, 881–896.

Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press.

van Lelyveld, I., & in 't Veld, D. (2014). Finding the core: Network structure in interbank markets. *Journal of Banking & Finance*, *49*, 27–40.

McAssey, M. P. (2013). An empirical goodness-of-fit test for multivariate distributions. *Journal of Applied Statistics*, *40*, 1120–1131.

McLachlan, G., & Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance*, *35*, 1114–1127.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical review E*, *66*, 016128.

Newman, M. E. (2010). *Networks: an introduction*. Oxford university press.

Newman, M. E., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, *68*, 036122.

Paltalidis, N., Gounopoulos, D., Kizys, R., & Koutelidakis, Y. (2015). Transmission channels of systemic risk and contagion in the European financial network. *Journal of Banking & Finance*, *61*, S36–S52.

Puliga, M., Caldarelli, G., & Battiston, S. (2014). Credit default swaps networks and systemic risk. *Scientific Reports*, *4*, 6822.

Schwarcz, S. L. (2008). Systemic risk. *Georgetown Law Journal*, *97*, 193–249.

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *The Journal of Finance*, *53*, 1821–1827.

Torri, G., Giacometti, R., & Paterlini, S. (2017). Sparse precision matrices for minimum variance portfolios. Available at SSRN 2965092.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.

Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports*, *3*, 2522.

Appendix A. List of Banks and Summary Statistics

Bank Name	**Short Name**	**Country**	**Area**	*mean*	*sd*	*sk*	*ku*	*v (d.f.)*	*ks t (1%)*	*ks norm. (1%)*	**ECB**	**EBA**
B.ca M. Dei Paschi Di Siena S.p.A.	M.te Paschi	Italy	S	0.00	0.09	0.26	3.87	6.47	not rejected	not rejected	Yes	Yes
Banco Popolare – So.Co.	Banco Popolare	Italy	S	0.00	0.09	0.63	4.48	5.39	not rejected	not rejected	Yes	Yes
Banca Pop. Di Milano – So. Co. R.L.	BPM	Italy	S	0.00	0.08	0.68	6.4	2.65	not rejected	rejected	Yes	No
BNL - Gruppo BNP Paribas	BNL	Italy	S	0.00	0.09	0.08	3.79	7.60	not rejected	not rejected	Yes**	Yes**
Intesa Sanpaolo S.p.A.	Intesa S.P.	Italy	S	0.00	0.11	0.30	3.65	8.59	not rejected	not rejected	Yes	Yes
Mediobanca S.p.A.	Mediobanca	Italy	S	0.00	0.08	0.03	6.33	2.50	not rejected	rejected	Yes	No
UniCredit S.p.A.	Unicredit	Italy	S	0.00	0.10	0.45	4.45	6.76	not rejected	not rejected	Yes	Yes
Banco Santander S.A.	B.co Santander	Spain	S	0.00	0.10	0.09	4.14	7.60	not rejected	not rejected	Yes	Yes
Banco Bilbao Vizcaya Argentaria S.A.	BBVA	Spain	S	0.00	0.09	0.09	3.90	8.20	not rejected	not rejected	Yes	Yes
Banco Comercial Português	Banco Com. Port.	Portugal	S	0.00	0.09	0.79	6.39	2.44	not rejected	rejected	Yes	No
BNP Paribas	BNP Paribas	France	C	0.00	0.10	0.04	3.95	6.38	not rejected	not rejected	Yes	Yes
Crédit Agricole S.A.	Cr. Agricole	France	C	0.00	0.09	0.35	3.90	8.71	not rejected	not rejected	Yes	Yes
Crédit Lyonnais	Credit Lyonnais	France	C	0.00	0.09	0.29	3.83	8.80	not rejected	not rejected	Yes***	Yes***
Société Générale S.A.	Soc. Gen.	France	C	0.00	0.09	0.37	4.04	6.89	not rejected	not rejected	Yes	Yes
KBC Bank	KBC Bank	Belgium	C	0.00	0.06	0.91	7.94	2.38	not rejected	rejected	Yes	Yes
Bayerische Landesbank	Bayerische LBank	Germany	C	0.00	0.06	0.28	6.89	1.75	rejected	rejected	Yes	Yes
commerzbank AG	Commerzbank	Germany	C	0.00	0.09	0.60	6.63	5.52	not rejected	not rejected	Yes	Yes
Deutsche Bank AG	DB	Germany	C	0.00	0.09	0.08	5.24	5.78	not rejected	not rejected	Yes	Yes
Erste Group Bank AG	Erste Group	Austria	C	0.00	0.07	0.28	8.37	2.41	not rejected	rejected	Yes	Yes
Coöperatieve Rabobank U.A.	Rabobank	Netherlands	C	0.00	0.07	-0.11	4.39	4.25	not rejected	not rejected	Yes	Yes
ING Bank	ING Bank	Netherlands	C	0.00	0.08	0.05	3.89	6.08	not rejected	not rejected	Yes	Yes
Skandinaviska Enskilda Banken	Skandinaviska	Sweden	O	0.00	0.05	0.36	6.47	2.15	not rejected	rejected	No*	Yes
Svenska Handelsbanken	Svenska Handb.	Sweden	O	0.00	0.06	0.81	6.48	2.88	not rejected	rejected	No*	Yes
Danske Bank	Danske Bank	Denmark	O	0.00	0.06	0.20	5.22	3.17	not rejected	not rejected	No*	Yes
Barclays Bank PLC	Barclays Bank	UK	O	0.00	0.10	0.30	4.82	7.21	not rejected	not rejected	No*	Yes
Bank of Scotland PLC	Bank of Scotland	UK	O	0.00	0.08	0.53	6.48	4.11	not rejected	not rejected	No*	No
HSBC Bank PLC	HSBC Bank	UK	O	0.00	0.09	0.11	4.35	6.59	not rejected	not rejected	No*	Yes
Lloyds Bank PLC	Lloyds Bank	UK	O	0.00	0.09	0.64	5.67	6.75	not rejected	not rejected	No*	Yes
Natwest Bank PLC	Natwest Bank	UK	O	0.00	0.08	0.47	5.49	4.86	not rejected	not rejected	No*	No
Royal Bank of Scotland PLC	RBS	UK	O	0.00	0.09	0.80	6.35	5.79	not rejected	not rejected	No*	Yes
Standard Chartered Bank	Std. Chartered	UK	O	0.00	0.09	0.14	4.08	5.72	not rejected	not rejected	No*	No

Table A.7: *Area* indicates the geographical location of the banks (S refers to Southern Europe, C to central Europe and O to countries outside Eurozone). Data reports mean, standard deviation (sd), skewness (sk) and kurtosis (ku) for the period 01/01/2009 – 30/06/2016. Degrees of freedom (v) refer to the univariate t distribution fitted on the data. *ks t (1%)* and *ks norm (1%)* refer to the Kolmogorov Smirnov test for goodness of fit of t and normal distribution with $\alpha = 0.01$ (the value refers to the rejection of the distribution considered). The last two columns refer to being under European Central Bank supervision, and of being in the group of financial institutions considered by the EBA EU-wide stress test exercise, respectively. *Not in Eurozone. **BNP Paribas Group. ***Crédit Agricole Group.

Appendix B. Network Analysis

The characteristics of a network can be summarized using a series of indicators. Such indicators are typically categorized into *local* and *global*, where local indicators refer to characteristics of individual nodes and global indicators refer to properties of the network as a whole.

In our analysis, we focus on undirected weighted networks (i.e., the edges identified by penalized partial correlation networks do not have a direction but only a weight corresponding to the estimated partial correlation). We are primarily interested in two features of the network: the centrality of the nodes and the presence of structural properties such as core-periphery or community structure.

Appendix B.1. Centrality Measures and Freeman Centralization

Centrality measures are indicators used to quantify the importance of individual nodes in the network. We present here four typical measures: *strength centrality*, *eigenvector centrality*, *Bonacich power centrality* and *betweenness centrality*.

The *strength centrality* of node i is defined as:

$$c_i^S = \sum_{j=1}^{m} w_{ij}, \tag{B.1}$$

where c_i^S is the *strength centrality*, or *strength* of node i, m is the total number of nodes and w_{ij} is the weight of edge ij. This measure can be considered an extension of *degree centrality* to weighted networks (Barrat et al., 2004).

The *eigenvector centrality* is a measure that assigns to each node a score based on the principle that connections to important nodes contribute more to centrality than connections with less central nodes, showing a recursive nature. Formally, the ith node's centrality score is proportional to the sum of the scores of all the nodes which are connected to it:

$$c_i^E = \frac{1}{\delta} \sum_{j=1}^{m} a_{ij} c_j^E, \tag{B.2}$$

where c_j^E is the *eigenvector centrality* of node j, a_{ij} is the ijth element of the adjacency matrix $\boldsymbol{A}$ and δ is a constant. Under suitable conditions, *eigenvector centrality* can be computed as the eigenvector corresponding to the largest eigenvalue of the adjacency matrix $\boldsymbol{A}$. *Eigenvector centrality* can be generalized to weighted networks using the weighted adjacency matrix $\boldsymbol{W}$ instead of $\boldsymbol{A}$.

Bonacich power centrality is another recursive centrality measure, defined as:

$$c_i^B = \sum_{j=1}^{m} (\alpha + \beta c_j^B) a_{ij} \tag{B.3}$$

where c_i^B is the *Bonacich power centrality* of node i, α and β are two constants, and a_{ij} is ijth element of the adjacency matrix $\boldsymbol{A}$ (Bonacich, 1987). Similarly to *eigenvector centrality*, a node's centrality is influenced by the neighbours' centralities. For convenience, we consider the case in which $\alpha = \beta = 1$.

Finally, we consider *betweenness centrality*, that is based on the concept of the shortest path (i.e., the uninterrupted chain of edges that connects any two nodes with the minimum number of steps). *Betweenness centrality* for each node is computed as the number of these shortest paths that pass through it (Newman, 2010).

Appendix B.2. Freeman Centralization

Network centralities are local measures that apply to individual nodes. It is also possible to consider global measures that characterize the centralization of the entire network computing the *Freeman centralization* measure as follows (Freeman, 1978):

$$C = \frac{\sum_{i=1}^{m} c_* - c_i}{\max\left(\sum_{i=1}^{m} c_* - c_i\right)}, \tag{B.4}$$

where c_i is a local centrality measure for the ith node, c_* is the centrality corresponding to the most central node in the network and $\max \sum_{i=1}^{m}(c_* - c_i)$ is a normalization factor computed on the network with the highest centralization index for a given number of nodes.[12] In other words, *Freeman centralization* calculates the difference between the centrality of the most central node and all the others, Therefore, a network with a high centralization has a node, or a group of nodes, that has a high centrality and a large number peripheral nodes.

Appendix B.3. Modularity

Modularity is a measure that allows us to measure how well a certain partition describes the network. A high level of *modularity* indicates that the number of edges connecting nodes in the same partition is higher than expected. Given a partition $G = \{G_1, \ldots, G_p\}$, we can define modularity Q as follows:

$$Q = \frac{1}{2M} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(w_{ij} - \frac{c_i^S c_j^S}{2M} \right) \delta(g_i, g_j), \tag{B.5}$$

where w_{ij} is an element of the weighted adjacency matrix, c_i^S is the strength of node i, $M = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}$, g_i is the group in the partition in which the element i belongs and $\delta(g_i, g_j)$ is 1 if $g_i = g_j$ and 0 otherwise.

Modularity can be computed on any partition of the network. In this paper, we use the extention of *modularity* suitable for signed networks proposed by Gómez et al. (2009).

Appendix B.4. Clustering Coefficient

The *clustering coefficient* measures the tendency of the network to create triangles. In a network with communities it is likely that, for a given node, two of its neighbouring nodes are connected to each other,

[12] For several centrality measures the most centralized network is star-shaped, while for the eigenvector centrality it is a disconnected network with a single edge connecting two nodes and a large number of disconnected nodes.

"closing the triangle".

We consider the *clustering coefficient* for weighted networks proposed by Fagiolo (2007), which is defined as:

$$c_i^w = \frac{(\boldsymbol{W}^{[1/3]})^3_{ii}}{k_i(k_i-1)}, \tag{B.6}$$

where k_i is the degree of node i and $\boldsymbol{W}^{[1/3]} = \{w_{i,j}^{1/3}\}$, that is, the matrix obtained by the weight matrix $\boldsymbol{W}$ by taking the third root of each entry. The measure takes values in [0,1] and is equal to the non weighted *clustering coefficient*, when the weights are either 0 or 1. The *clustering coefficient* c_i^w is a local measure that applies to each individual node i. Still, it is possible to derive a global clustering measure C^w computed as the average of c_i^w over all the nodes. Because the *clustering coefficient* is not defined for negative weights, we set the weight for negative edges to zero.

Appendix B.5. Assortativity

An useful indicator to describe the structure of a network is *assortativity*, defined as the Pearson correlation coefficient for a nodes' characteristic computed over the edges of the network. This indicator measures the extent to which nodes with similar properties are connected to each other and can be computed for virtually any property assigned to the nodes, including categorical ones (Newman, 2010). Typically it computed using the *degree centralities* of the nodes. In a weighted network, the expression for *assortativity* with respect to a scalar characteristic is the following:

$$r = \frac{\sum_{k=1}^m w_k(x_k-\mu_x)(y_k-\mu_y)}{\sqrt{\sum_{k=1}^m w_k(x_k-\mu_x)^2 \sum_k w_k(y_k-\mu_y)^2}}, \tag{B.7}$$

where k is the index corresponding to each edge, w_k is the weight of edge k, x_k and y_k are scalar characteristics of the originating and destination nodes (in the paper we consider centrality measures) and μ_x and μ_y are the respective average values weigthed by the edges' weights. In other words, the *assortativity* is the weighted correlation coefficient over the edges of each network and thus lies in the range [-1,1] with $r = 1$ indicating perfect *assortativity* and $r = -1$ indicating perfect *disassortativity*. Note that *assortativity* is not defined for negative weights, therefore in the computation of the indicator we set the weight for negative edges to zero.

An assortative network can be consistent with a core-periphery structure, where we observe a core of highly central nodes surrounded by a less dense periphery of nodes with lower centrality. Newman & Park (2003) show that a high level of *assortativity* can be found also in network characterized by the presence of communities, typically social networks.

Appendix C. Robustness Checks

Here, we report some robustness checks on the model by testing different specifications for the input data. We control for two potential sources of misspecification in the model (ARMA-GARCH effects in

the time series and exposures of the banks to common risk factors), and we investigate the estimated network structure computed by considering only distressed market periods, which are characterized by increasing CDS spreads. We report a set of network indicators computed on the estimated networks in Table C.8. For brevity, we present only the data for the *during-crisis* period.

First, we control for the presence of serial correlation and heteroskedasticity. The *tlasso* model requires the data to be approximately i.i.d., and the presence of serial correlation in CDS log-differences or in their variance might be a problem for the estimation. We consider two different multivariate models: first a Constant Conditional Correlation (CCC) GARCH(1,1) model (Bollerslev, 1990), and a Dynamic Conditional Correlation (DCC) GARCH(1,1) (Engle, 2002). For both models, we estimate an ARMA(1,1) process for the mean. Once the model is fitted, we estimate the *tlasso* model on the residuals, in a two-step procedure similar to Anufriev & Panchenko (2015) and Billio et al. (2012). We report some of the indicators of Section 4.3 computed on the residuals of the GARCH models in Table C.8 (columns b,c). The results are rather similar to the ones computed directly on the log-differences of CDS spreads (column a), consistently with the fact that volatility clustering and serial correlation in the data were not particularly relevant.

As a second robustness check, we apply a factor model to the data to control for the exposure to conditioning variables. The factors we consider are 5-year sovereign CDS spreads of the 11 European countries to which the banks in our sample belong (i.e., Austria, Belgium, Denmark, France, Germany, Italy, Netherland, Portugal, Spain, Sweden and United Kingdom). In order to avoid collinearity issues, we perform a principal component analysis, retaining the first 8 orthogonal factors, which together explain more than 80% of the variability in the dataset. From Table C.8, (column d), we see that the value of the indicators is rather similar to those computed from original data. Notably, the modularity of the country partition is still positive, indicating that the clustering of banks along national borders is present even after controlling for the exposure to sovereign debt.

Finally, we test if the network structure is different in presence of market distress. We focus on the periods characterized by upward movements of the CDS spreads, that is, increases in credit risk. We use the following method to identify these periods of distress. We first construct an equally weighted porfolio of bank CDS spreads and then select the points in time at which the log-returns of the porfolio are positive, corresponding to an average increase of the CDS spreads. In this way, we obtain for the *during crisis* period 102 observations, which we use to compute the sparse partial correlation network with *tlasso*. Also in this case, the values of the network indicators presented in Table C.8 (column e) are similar to those obtained for the original specification, indicating that the partial correlation networks are stable also in periods of increasing spreads. This result is consistent with the fact that the distribution of the CDS spreads log-difference is symmetric, as implied by the multivariate t-Student assumption and suggested by the analysis of the data in Section 4.1.

Table C.8: Robustness check - Network indicators

	During-crisis period				
	Raw data (a)	CCC-GARCH (b)	DCC-GARCH (c)	Country factors (d)	Distressed market (e)
Density	49.25%	47.53%	47.74%	44.95%	43.44%
Avg strength	0.95	0.95	0.95	0.93	0.96
Largest connected component	100%	100%	100%	100%	100%
Positive edges %	87.77%	90.95%	90.99%	90.43%	84.65
Positive edges % (weighted)	96.59%	97.43%	97.43%	97.63%	94.91
Freeman Centralization	0.47	0.43	0.43	0.48	0.47
Coreness	0.11	0.12	0.12	0.13	0.11
Modularity (wrt country)	0.36	0.34	0.34	0.36	0.37
Clustering coefficient	2.70%	2.54%	2.54%	2.60%	2.82%
Assortativity (wrt centrality)	0.61	0.59	0.59	0.66	0.43

Network indicators computed on *tlasso networks* in the *during-crisis* period (01/01/2009 – 31/12/2012) with alternative model specifications. *Raw data (a)* refers to the network computed with the specifications presented in Section 4. *CCC-GARCH (b) and DCC-GARCH(c)* refer to the network computed on the residuals of Constant and Dynamic Conditional correlation GARCH model fitted on the log-differences of CDS spreads, respectively. *Country factors (d)* refers to the network computed on the residuals of a factor regression where the factors are Sovereign European CDS spreads for 11 European countries. *Distress market (e)* refers to the network computed in periods of increasing spreads.