IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"

14-17 February 2022

A multi-layer dynamic network for significant European banking groups[1]

Annalaura Ianiro and Joerg Reddig,
European Central Bank

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

A multi-layer dynamic network for significant European banking groups

Ianiro, Annalaura; Reddig, Jörg[1]

Abstract

In the aftermath of the financial crisis there was a clear demand for granular statistical data which would allow user to gain insights into topics ranging from financial stability, monetary policy, banking supervision and diverse research questions which could not be answered before. The statistical function of the Eurosystem answered this demand by collecting new and highly granular data sets on securities, money market transactions, derivative trades, and loans which are accompanied by a collection of master data on institutions allowing interlinking this information. In this work, we first show how we integrate these data to build a multi-layered network of the significant European banking groups. Then, we proceed to analyse the data with a new Python toolkit, NATkit, specifically developed to work with multi-layer dynamic networks.

Keywords: Multi-layer dynamic network, significant banking groups, granular data

JEL classification: G21

The views expressed are those of the authors only and do not necessarily reflect those of the European Central Bank.

[1] Annalaura Ianiro, Annalaura.Ianiro@ecb.europa.eu, European Central Bank

Jörg Reddig, Joerg.Reddig@ecb.europa.eu, European Central Bank

Contents

1. Introduction

After the global financial crisis, among the many lessons learned, policymakers got reminded that for non-linear events driven by financial instability a timely analysis of interconnected granular data can help to better understand the developments in the financial sector at the level of individual agents and their interactions in various markets, helping to prevent or contain financial crises. Consequently, the European Central Bank (ECB) invested in appropriate modelling tools and started to purposely collect more granular data, in order to support the creation of multipurpose interconnected data sets that allow the insights that have been missing before. Given the way they were designed and conceived, these datasets have the potential to be a game changer for understanding the landscape and interconnectedness of the euro area financial markets. It is, thus, important to analyse interconnected markets in a joint manner. Only then, we might be able to uncover meaningful insights –i.e. exposure of financial institutions to certain economics sector, concentration risk, substitution effects between asset classes, etc. – and the potential propagation channels of shocks.

To do so in a systematic and coherent way, granular datasets need to be integrated with each other and be analysed jointly. Indeed, as former ECB President Mario Draghi stated, "developing the analytical toolkit to adequately monitor interconnectedness and contagion requires granular datasets, and the ability to map and link data across entities and markets" (Draghi, 2019).

The ECB's Data Committee sponsored a project to develop tools that would allow users from various business areas and policy makers to apply scalable and reproducible advanced analytics on the interconnections of institutions and financial markets via the use of integrated granular data. These would include loan-by-loan and security-by-security (issues and holdings) data, money market transactions, and over-the-counter derivatives transactions. Such data could also be combined with supervisory information on banks like COREP, FINREP and other supervisory data, monetary policy data on MFI balance sheets and interest rates, monetary policy operations as well as commercial datasets like Orbis/Bankfocus on firms and financial institutions. Advanced analytics based on the resulting integrated granular dataset would allow supplementing the standard models based on macroeconomics variables, in particular by being able to better capture non-linear contagion effects that in particular stem from the interconnectedness of institutions in different financial markets.

Advanced, scalable and reproducible analytics not only need common data, but also common analytical tools that can be easy to use, well document and extensible. The integrated granular dataset mentioned above can be seen as a multi-layer dynamic network, in other words a combination of multiple individual networks (the layers), that change over time (dynamic). The available tools for network analysis lack the functionalities needed to deal with this type of networks, forcing analysts either to come up with custom solutions (which are very specific and difficult to scale) or to analyse the network as if not multi-layer nor dynamic (which loses important information). NATkit (Network Analytics Toolkit) is an analytical toolkit that was developed (as a Python library) to fill in this gap and analyse multi-layer dynamic networks with the proper framework. Its purpose is to provide an easy-to-use toolset, which could help to explore and understand complex multi-layer dynamic network data. The toolkit leverages on existing visualisation and network analytics libraries, building upon them the framework and functions to properly handle this type of networks.

This work is a showcase of the tool's functionalities on a sample of the multi-layer network of significant institutions of the euro area, built combining granular data loans, securities, derivative and money market transactions. It is structured as follows: section 2 offers a brief overview of the literature on networks analytics for multi-layer networks, with specific focus on financial networks. It explains why multi-layer networks need a specific analytical approach and provides a formal definition of what a multi-layer network is. Section 3 describes the dataset that are used for building the individual layers of the network while section 4 describes how exactly the data were integrated and the individual layers were constructed with an example of the securities layer. Information about the toolkit and its empirical application to the multi-layer network are presented in section 5. The final section concludes and gives an outlook on the future work.

2. Related Work – The Approach for Multi-Layer Networks

Many complex real-world phenomena can be represented as a network. Mathematically, a network is a *graph*, in other words a collection of vertices (nodes) connected among each other by lines (edges), making the study of complex networks the domain of graph theory.
In the beginning, complex networks with no apparent structure were described with *random graphs*. Random graphs, also known as *Erdős-Rényi graphs* from the names of the two mathematicians who first theorised them in 1950, have a fixed set of nodes, N, while the existence of an edge between any two nodes is equal to a probability p. The beauty of this model lies in its simplicity and nice properties; however, its ability to properly represent complex network systems has been since reconsidered: the topology of many real-world networks can hardly be described as random. Therefore, many tools and measures have been developed to properly describe the topologies of real-world graphs. Réka and Barabási (2002) provide an interesting overview on the statistical mechanics of complex networks, still a relevant read despite its age.
The specific literature of financial applications of network analysis began with the seminal contributions by Allen and Gale (2000) and Freixas et al. (2000) that recognised the importance of the structure of interconnections between financial institutions in a theoretical framework. From these studies, a rich literature has developed, defining the canons and standards of network analysis for complex financial and interbank networks. Such standards, as argued by Battiston et al. (2014) and Kurant and Thiran (2006), often ignore whether a network is multi-layer or not and can be divided into two types. One type studies the graph resulting from the aggregation of all the edges between a certain set of nodes, regardless of the type of relationship they represent. The other type treats the different layers of the same network as separate entities. Hence, despite their much more complex reality, for a long time interbank and financial networks have either been compressed into a single aggregate layer or different layers have been studied in isolation. Both these approaches overlook important information and only paint an incomplete picture about the structure and functioning of financial relationships among market players.

Representing the network as a *multi-layer graph* or *multiplex* allows to overcome the limitations of the standard approach. When analysing the financial and thus interbank networks, in particular, the literature has highlighted at least three advantages of the multi-layer approach:

- It accounts for the heterogeneity of each layer. Bargigli et al (2015) finds that several topological and metric properties are layer-specific, whereas other properties are of a more universal nature.

- It allows to capture non-linearity, which is important for risk assessment. Poledna et al. (2015) show the existence of non-linearity in the way risks are aggregated: in their application, systemic risk for the aggregated network may be considerably larger than for the sum of the component sub-networks.

- It can provide relevant metrics for banking regulators and supervisors. Aldasoro and Alves (2015) propose a measure of systemic importance that allows to decompose the global systemic importance index for any bank into the contributions of each of the sub-layers. Such a measure can lead to better tailored policy responses.

Having briefly highlighted the advantages of the multi-layer approach, we will explain what a multi-layer network, or multiplex, is, as well as extend the definition to include the dynamic component. While Kivela et al (2014) provide a very extensive formalisation of multi-layer networks, for the purpose of this work we will consider the definition given in Battiston et al (2014), which is more intuitive. According to them, a multiplex is "a network where each node appears in a set of different layers, and each layer describes all the edges of a given type". A dynamic multiplex is, therefore, the set of the time-specific multiplexes over the given time interval.
More formally, we could think of a network in terms of its *adjacency matrix*. Focusing on a single time period t, an adjacency matrix is a square nxn matrix, where n is the number of nodes in the network. Therefore, an entry e_{ij} represents an edge connecting node i to node j. When the network is unweighted, such entry is equal to 1 if an edge exists, 0 otherwise. In case of a weighted network, e_{ij} is equal to w_{ij} (the weight value) if there exists an edge, 0 otherwise. Another feature of the network is whether it is directed or undirected[2]. If the network is undirected, the adjacency matrix is symmetric, so that $e_{ij} = e_{ji}$. Each layer of the multi-layer network has its own adjacency matrix. Overlapping them (summing their entries together point by point) allows to obtain the *multiplex adjacency matrices*. Indeed, there are two types (three in case of weighted networks) of multiplex adjacency matrices for each multi-layer network. These are:

- A: aggregated multiplex adjacency matrix. This matrix has entry a_{ij} equal to 1 if there exists an edge in at least one of the layers, 0 otherwise.

- O: overlapping multiplex adjacency matrix. This matrix is built by overlapping the unweighted adjacency matrices of all the layers. Therefore, it has entry

[2] Financial networks are usually directed. For example, in case of loan exposures the two counterparties have distinct roles: one is the lender, the other is the borrower. An undirected network is, for example, friendship on Facebook: if person A is friend of person B, the opposite is always true. When, instead, we look at the network of Twitter following, which is directed, if person A follows person B, this doesn't imply that person B follows person A.

$o_{ij} = \sum e_{ij}$. In case there is no edge between nodes i and j in any of the layers, the entry is 0.

- O^w: weighted overlapping multiplex adjacency matrix (only for weighted networks). This matrix is built by overlapping the weighted adjacency matrices of all the layers. Therefore, it has entry $o_{ij}^w = \sum w_{ij}$. In case there is no weighted edge between nodes i and j in any of the layers, the entry is 0.

A dynamic multi-layer network is represented by the set of the above matrices for each specific time period t in consideration.
Each of the matrices can be viewed and analysed as its own network. However, in order to properly account for the non-linear relationships that might exist among the different layers, it is important to consider both the multiplex representations and the separate layers in the analysis. Section 5 further elaborate and explain this concept.

3. The datasets

Having defined what a multi-layer dynamic network is, we now turn to describing which granular statistical data are available at the ECB that we used for building the multi-layer network. All these datasets are stored on the ECB's Data Intelligence and Service Center (DISC) platform. This is the ECB's Hadoop-based datalake that allows to query the data via SQL or use them directly on an Apache Spark cluster.

3.1. List of significant banking groups (ROSSI list)

Starting point for the creation of our multi-layer dynamic network for banking groups is to define which entities shall be included in the sample. As almost all available granular datasets comprise information on the largest banking groups in the Euro Area and participating member states which are directly supervised by the Single Supervisory Mechanism (SSM), we decided to use this list as the sample for our network.
The ECB maintains a list of all significant banks under its direct supervision and less significant banks under its indirect supervision, which is publicly available on the SSM homepage.[3] As of 1 November 2021, this list of significant institutions comprised a total of 115 banking groups. However, the list is reviewed every year and ad-hoc assessments are carried out during the year whenever necessary to assess the significance status of banking groups based on a size criterion which considers the total value of the supervised entity's or the supervised group's assets, at consolidated level.[4] Therefore, the number of supervised institutions changes over time.
Internally, this list is made available via the Repository of the SSM Supervised Institutions (ROSSI). From this list of banking groups, we extract the names of the significant head institutions under direct SSM supervision as well as their Legal Entity Identifiers (LEI), which is a unique global identifier for legal entities participating in

[3] See also https://www.bankingsupervision.europa.eu/banking/list/html/index.en.html.

[4] A detailed description on the assessment of the significance status can be found here.

financial transactions. These identifiers are used to connect the supervisory entities with the statistical datasets.

3.2. Group structure of the significant banking groups (RIAD)

Once we have identified the list of significant institutions relevant for our network, the next step is to get data on the group structures of these banking groups. This way, we will get a view on all entities that belong to the group, which will be the foundation of the construction of the network.

This information can be retrieved from the Register of institutions and affiliates database (RIAD), the Eurosystem's unique master data repository. It not only contains detailed information on financial institutions but also on their group structures. In fact, four types of group structures are available for different purposes. In the analysis here, we use the RIAD group structure type A, which is based on all *direct and indirect control relationships* in a group. RIAD data are constantly updated and enriched. For the purpose of this analysis, we use monthly snapshots on the end-of-month values for the group structures to match the other datasets which we use to construct the different layers of our network.

3.3. Securities layer (CSDB / SHSG)

For the securities layer, we employ two granular datasets: data on the issuance activities of the banking groups, which we can retrieve from the Centralised Securities Database (CSDB), and the holdings of securities, which are available in the Securities Holdings Statistics Database by Group dataset (SHSG).

The CSDB is a reference database for securities which aims to cover all securities relevant for statistical purposes of the European System of Central Banks. It comprises information on debt securities, equity instruments and investment fund shares which are stored on a security-by-security basis where each instrument is identifiable by the International Securities Identification Number (ISIN). For each of these securities a vast number of attributes are available. For the purposes of building a securities layer in the multilayer network the CSDB gives us important information on the issuers, like issuer name, identification number, and sector classification and on the securities, like instrument type, prices, total issuance, original and remaining maturity, and ratings. The data in the CSDB are available on the ECB's data lake in monthly frequency and cover the periods from starting from 2014.

Sources for this database are commercial data which are bought from several providers as well as input from the Eurosystem central banks. Together, the ECB and the National Central Banks do a constant quality assessment on the data. This guarantees the necessary coverage of securities information used in various statistics of the Eurosystem and a high-quality standard to make the data fit for purpose.[5]

While the CSDB allows us a very good view on the issuance of securities by the significant banking groups, the SHSG dataset provides the reported holdings of securities by the banks. As of 2018-Q3 the significant banking groups must report their holdings of debt securities and equity instruments (including investment funds

[5] For more information on the CSDB, please see *The "Centralised Securities Database" in brief* (European Central Bank, 2010).

shares) on an entity-by-entity basis. This means, that all securities holdings must be broken down by the entities of the group that are holding theses securities on a global level. Thus, we get a complete overview which part of the banking group is holding the securities, inside and outside the euro area.
Like CSDB data, the amounts of securities held by the significant institutions are reported on a securities-by-securities level and can be identified by the ISIN code. This allows us as well to merge the two granular securities datasets and construct the securities layer of the multilayer network, where the so-called cross holdings between the banking groups form the edges of the network.

3.4. Loans layer (AnaCredit)

The layer describing the loan exposures between the significant banking groups is based on the AnaCredit dataset. The dataset contains loan-by-loan information collected from euro area banks extended to corporations. These data are reported at monthly frequency and are available as of September 2018. They comprise more than 80 attributes which give a detailed picture of the nature of these loans. Among others, information on the outstanding amount, maturity, interest rate, collateral/guarantee, and on counterparties are collected for each of the individual loans. The data are collected for each entity of a banking group separately (unconsolidated), which needs to be considered for the calculation of the total loans of a banking group.
This makes AnaCredit an incredibly rich dataset that can be used for various analytical purposes. However, as more than 25 million credit instruments are recorded every month and the large number of attributes, this dataset is also computationally very resource demanding.

3.5. Money market transactions data (MMSR)

The money market statistical reporting (MMSR) data are collected on transaction-by-transaction basis from a sample of euro area reporting agents. This way, it covers the most relevant institutions which are active on the European money market. It provides information on the secured, unsecured, foreign exchange swap and overnight index swap euro money market segments. The ECB uses this information to calculate euro short-term rate (€STR).
For the money market layer, we therefore have two important differences: Firstly, the data collection only comprises a sample of the money market participants. This means that the network does not cover all significant institutions but only a subset of them. An additional complication is that often trades between two market participants are conducted via a central counter party. In those cases, we employ a matching procedure to identify the indirect trading partners. Secondly, the MMSR dataset focuses on daily short-term financial transactions between the money market participant. In contrast, the securities and loan layers comprise end-of-period stock data that are reflected in the balance sheets of the reporting institutions.
The granular MMSR trade data include amongst other attributes the interest rate for the transaction, the volume and counterparty information as well as collateral type information. Moreover, both sides of a transaction (lending and borrowing) are reported by the parties involved. In our analysis of the MMSR data we focus on the secured money market segment as this is the most active part of the market. Analysing these data can bring very valuable information on the structure of the

money market. Moreover, it helps to identify liquidity shortages on the market and thus helps to monitor the risks for market participants.

3.6. Derivative data layer (EMIR)

European Market Infrastructure Regulation (EMIR) dataset contains a vast amount of data on derivatives markets and is collected via authorised trade repositories by the European Securities and Markets Authority (ESMA). The EMIR regulation covers transactional-level derivatives data for all counterparties established in the euro area as well as all contracts where at least one entity is located within the euro area or where the reference obligation is sovereign debt of a euro area member.
The data are reported at daily frequency and cover 129 data attributes which are reported for each contract. On the one hand, these attributes cover information on the counterparties involved and on the other information on the characteristics of the contract (for example the type of derivative, the underlying, outstanding amounts and prices), as well as details about how the contracts were executed. This wealth of information makes it one of the largest datasets currently available at the ECB.

3.7. Other datasets used

Apart from these granular and master data, we also used some bank balance sheet information from the supervisory FINREP data and the statistical Balance sheet items statistics. We mostly used data on the total balance sheets of the banking groups to enrich the information on the nodes in the network.
An overview of the used granular information is shown in table 1, which shows as of when data are available on the ECB's datalake as well as the frequency that these datasets have.

Granular datasets overview

Table 1

	Time frame	Frequency
Datasets		
AnaCredit	2018-09	Monthly data
EMIR	2014-02	Daily data
MMSR	2018-01	Daily data
CSDB	2014-01	Monthly
SHSG	2018-Q3	Quarterly data

Sources: ECB

4. Building the multi-layer dynamic network

The integration framework described in this section is the core building block of the multi-layer dynamic network of the significant institutions of the euro area. We would like to highlight that this framework is only one of the possible ways of integrating the datasets; in particular, it is an integration as end-users of each datasets, rather than an integration at the source. Trying to merge datasets that were not conceived for that implied considerable challenges, which we overcame to the best of our abilities. Despite its limitations, we hope that this work can become the basis of future analysis proving the value that integrating granular datasets have for an institution like the ECB, thus paving the way for data collection policies that envisage integration at the source.
One final remark: even if not explicitly written to keep the explanation easier to read, it's important to clarify that each dataset is joined for each reference period considered (i.e. the reference date filed of each dataset is used as a secondary key for each join).

4.1. Retrieving the group structure of the significant institutions

The construction of each layer of the multi-layer dynamic network starts with the integration of all the datasets together. We begin with the list of significant institutions that we got from the ROSSI database. As described in the previous chapter, this list is maintained by the supervisory function of the ECB. It includes two types of identifiers that can help to uniquely identify the parent company of the SSM supervised institutions: the LEI and the ESCB-internal RIAD code. While the LEI is an internationally agree-upon standard, it can happen that some entities do not have such a code or that there are doubts whether the correct code is stored in the database. Therefore, we choose the RIAD code, which is universally used for statistical dataset collected by the ECB, as our entity identifier to connect the datasets with each other.
Having the identifiers of the significant institutions, we proceed to retrieve the list of all the subsidiaries belonging to a specific banking group. To achieve this, we join the ROSSI list with the RIAD database to retrieve the group structure of each significant institution. As previously explained in section 3, RIAD provides four different types of group structures. We choose RIAD group structure type A, which includes all directly and indirectly control entities in a group[6]. Given our focus on the core banking group, we apply a filter to the join, so that only credit institutions in the group are taken into account. The group structure thus obtained will allow us to get an overview of the activities of all group members in each layer of the network. In the following sub-section, we provide an in-depth example on how, starting from the group structure, we are able to get the cross-holdings information of each banking group (i.e. who is issuing a specific security and who is holding it). We apply a similar methodology to the other datasets to build the layers of the network for loan data, derivatives data and the money market layer, but for brevity's sake, we will not describe them. It is important to highlight, though, that the different layers are constructed in a consistent way, as we start from the same group structure information.

[6] Specifically, it considers ownership relationships with an equity share greater than 50% and/or existence of control.

4.2 In depth example: building the securities cross-holdings layer

The first building block of the securities cross-holdings layer is retrieving the total issuance of each credit institution in each banking group. In practice, this implies joining all RIAD codes[7] identifying the group members with the CSDB, the master database for securities. This gives us a comprehensive list of securities issued by all the members of the banking group, where each security can be uniquely identified by its ISIN code. With it, it is also possible to retrieve attributes on the securities that we need for further analysis of the securities layer. For example, the CSDB contains information on total amount issued for each security, their price, the date of issuance, the original and remaining maturity, and the classification of the securities into debt or equity instruments. For the purpose of this work, we treat the securities layer one as one layer, aggregating all instruments at parent entity level. However, depending on the analysis, it could make sense to define additional sub-layers, e.g. the available attributes let us distinguish between short-term and long-term debt securities layers or between equity and debt layers.
The next step is looking at the holding side and retrieve the total holdings of each credit institution in each banking group. We collect this information by joining SHSG with the group structure, through the RIAD codes. This gives us a comprehensive list of securities held by all the members of the banking group, where each security can be uniquely identified by its ISIN code, as it is the case for the issuance side. Lastly, we need to combine the issuance and holding side, using the ISIN code as the unique identifier and main key for the join.

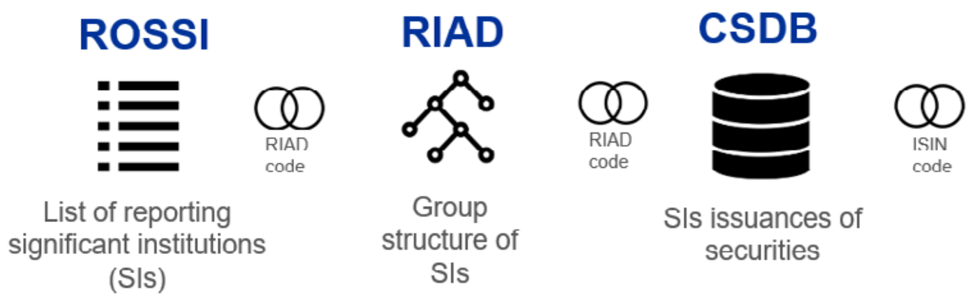


Figure 1: Stylised joins used to create the securities layer of the network

Figure 1 shows a stylised representation of the different joins that are used to create the securities layer for the multi-layer network. As we explained, there are four different datasets involved and two main different identifiers to merge the data. As a result, we get a comprehensive overview of the cross-holdings of securities among the significant institutions supervised by the SSM. Figure 2 shows a representation of this layer through a chord diagram. Thanks to the granular nature of the data used in this exercise and the rich set of attributes available, we are able, among others, to see from which countries the significant institutions are from (represented with the different colours in the diagram). Furthermore, other attributes can be used in a flexible way depending on the question that researchers want to tackle with the network.

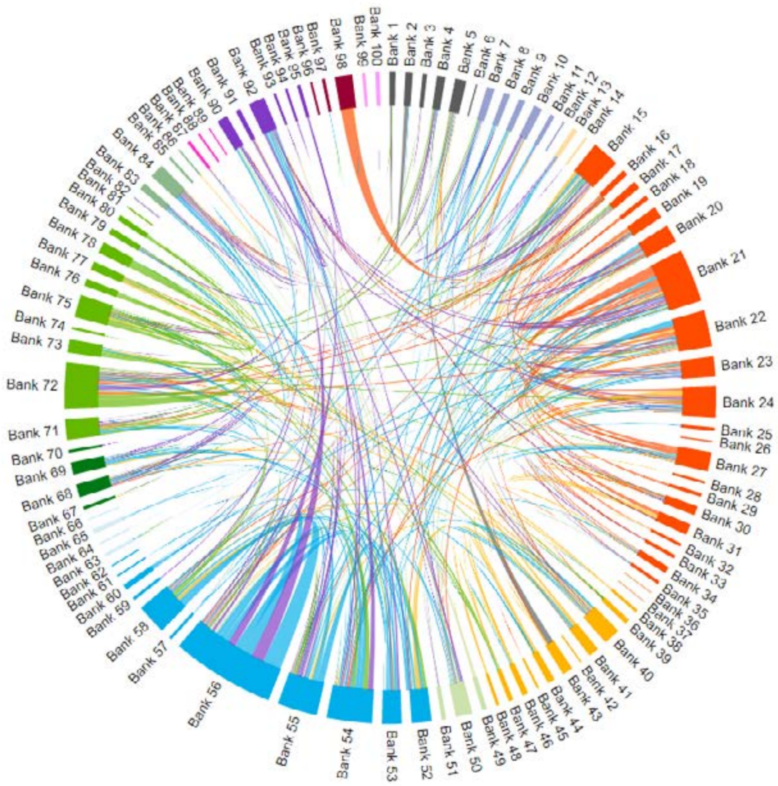


Figure 2: Chord diagram of the securities cross-holdings layer for period 2020-06.

4.3. The infrastructure

The integration framework is a set of SQL and PySpark scripts and their accompanying documentation, available to all colleagues at the ECB. Each script is built in blocks, prioritising flexibility and reusability. This allows us to easily update the code when necessary, as well as quickly adapt it to integrate new layers, as they become available. It is also possible to modify some of its parts to adapt it to different purposes (e.g. one might want to look at all entities belonging to a certain banking group, instead of just at the credit institutions).
Running the scripts allows the users to retrieve the integrated data, which can then be analysed with the preferred language (e.g. R, Python, MATLAB, etc.).

5. Analysing the multi-layer dynamic network

5.1 NATkit: An Overview

NATkit is a Python tool developed to analyse multi-layer dynamic networks. The existing standard Python libraries for network analyses, such as Networkx, lack the functions to deal with a multi-layered graph object; therefore, NATkit aims to fill this gap, providing a tool built specifically for the type of financial data analysed at the ECB. The first prototype of the tool has working versions of the two core modules, visualisation and topology.

5.1.1. Visualization module

This core module allows the users to visualise the network in an interactive way. Combining different Python visualization libraries, it provides easy-to-use functions to plot complex graph objects. The interactivity allows the users to interact with the plot itself, for example to get information on mouse-hoover, as well as to see the evolution of the network over time. As an extra feature, users have the possibility to use an edge bundling algorithm to better disentangle the edges in the graph and obtain a clearer view of the network.

5.1.2. Topology module

This core module enables the users to analyse a multi-layer network in the proper framework, thanks to a combination of visualizations and topology metrics, which have been updated from the traditional ones to account for the multiplex nature of the graph object. In particular, while some topology metrics can be computed on the aggregated and overlapping representations of the network, some others have to rely on a combination of the different individual layers.

5.2. Using NATkit to analyse the multi-layer dynamic network of significant institutions

NATkit can help researchers to deep-dive into complex network structures and explain the interactions between the different layers. The purpose of the remainder of this paper is to demonstrate on real data how to use NATkit in order to analyse a dynamic multi-layer network. Therefore, the following analysis is going to be atypical, in the sense that it is not driven by a specific research question defined by business users, but is rather motivated by its ability to showcase the capabilities of the toolkit. In its development, priority has been given to functionalities that can allow the users to perform an analysis of multi-layer and dynamic networks as comprehensive as possible.

5.2.1. The Layers

Recalling section 4, the integration of the different datasets has resulted into the following layers:

- **Securities cross-holdings layer**: this layer is the combination of ROSSI, RIAD, CSDB and SHSG. It contains information about the securities holdings of the significant institutions of the euro area.

- **Loan exposures layer**: this layer integrates data from ROSSI, RIAD and AnaCredit. It contains information of the loan exposures of euro area significant institutions towards each other.

- **Secured money market transactions layer**: this layer is obtained by joining ROSSI, RIAD and MMSR. It contains information on the secured money market transactions of the significant institutions of the euro area.

- **Interest rate derivative exposures layer**: this layer is built by joining ROSSI, RIAD and EMIR. It contains information on the interest rate derivative exposures of the significant institutions of the euro area towards each other.

All layers have information aggregated at parent company level and have been enriched with total asset information coming from FINREP. For the purpose of this analysis, a sample of the above data has been extracted, covering monthly snapshots[8] of the sub-layers from December 2019 to May 2020.

5.2.2. Building the multiplex layers

Building a multi-layer network can be a rather complicated endeavour since the very beginning, even in choosing which layers to use. We decided to use the layers aggregated as described in the previous sub-section; however, each of them has its own sub-layers[9], which could be aggregated into different multiplex layers. Therefore, for future analysis, it could be worth to discuss in-depth which sub-layers to consider and whether to include sub-sub-layers.
Of course, having the data for each single layer is not enough to build the multiplex layers. Some data reshaping is necessary and the toolkit provides a framework and related functions to do so.
After using the integration framework as described in section 4, each layer comes in a table form as an edge-list[10], with different naming conventions and fields. In order to build the multiplex layers, the different layers need to share the same structure and the same attributes. This means identifying:

- **Source nodes and target nodes:** given that each layer is a directed network, this first identification is crucial because it sets the basis for the economic interpretation of the connection. In this specific case, in each layer the source node is identified as the institution providing funds; therefore, each edge represents the exposure of such entity towards the target entity, which receives funds instead.

- **Edge weight**: in our data, this is the amount of the exposure. Specifically, it is the creditor's outstanding nominal amount for AnaCredit, the transaction

[8] While all other layers have source data at monthly or even daily frequency, the security cross holdings layer has SHSG as one of the sources, which has quarterly frequency for reporting. FINREP data are reported quarterly as well. In order to match the monthly frequency of the other layers, data as of Q4 2019 (December) has been used for January and February 2020; data as of Q1 2020 (March) has been used for April and May 2020.

[9] For example, the securities cross-holdings layer can be further divided into short-term securities versus long-term ones.

[10] An edge-list is a data structure used to represent a network as a list of its edges.

nominal amount for MMSR, the nominal value for the security cross-holdings and the notional amount for derivatives. All these amounts have been standardised by the total assets of the source node institution and have been rescaled between 0 and 1, in order to be aggregated in the multiplex layers.

- **Nodes attributes**: these are all the attributes of source and target nodes, which can be interesting for the analysis. In this specific case, the attributes are total assets and home country of the institution (home or host approach?).

- **Layer**: a field specifying the layer. The naming convention used is *ana* for AnaCredit, *mmsr* for the money market layer, *ch* for the security cross-holdings layer and *emir* for the derivatives layer.

- **Date**: a field specifying the time period (in this case the month) the data refer to.

Once all the layers follow this structure, they can be joined in the same table and pivoted, so that for each time period, there is the same set of source and target nodes in each layer.
At this point, we have the weighted edge-list for each layer. The next step is computing the unweighted edge-list, simply dividing each weighted edge-list by itself. Finally, it is possible to construct the following multi-layer edge-lists:

- **multi_a**: this is the aggregated multiplex layer, A. It is such that the weight of an edge is equal to 1 if there exists an edge in one of the sub-layers or 0 otherwise.

- **multi_o**: this is the overlapping multiplex layer, O. It is obtained by summing the unweighted edge-lists of each sub-layer. Therefore, given that there are four sub-layers, its values are integer and range between 0 and 4.

- **multi_ow**: this is the weighted overlapping multiplex layer, O^w. It is obtained by summing the weighted edge-lists of each sub-layer. Given the weight rescaling of each sub-layer, its values range between 0 and 4.

The reader might wonder why constructing three different multiplex layers. The weighted overlapping layer, multi_ow, characterises the edges from the perspective of the amount of the exposure. The bigger the weight in each layer and the more layers the edge exists in, the bigger the weight in the multiplex. This perspective, however, contains two others: the perspective that there exists a relationship in any of the sub-layers (a pure unweighted perspective) and the one that this relationship exists in one or more specific sub-layers. These other two aspects are captured respectively by the aggregated multiplex layer (multi_a) and the overlapping multiplex layer (multi_o). These different multiplex layers, as well as the weighted and unweighted versions of the four sub-layers, are equally important to consider in the analysis to better describe and understand the network as a whole.
The last step is transforming the edge-lists into proper graph objects. At this point, we can fully exploit NATkit's modules.

5.2.3. Visualizing the network

"Classic" network visualization is a powerful analytical tool: if done properly it can help gaining useful insights which can guide the analysis. There is one caveat though: when looking at a multi-layer network, it might not be that informative to visualise the layers all together. One layer is likely to be "messy enough" on its own. Therefore, when developing the visualization module, I tried to build a tool that could help users find a pattern in the chaos, focusing on one layer at a time.

Heatmaps are an underestimated way to visualise a network. They provide clear "fingerprints" of how a network looks like: how dense it is, if there are clusters, etc.

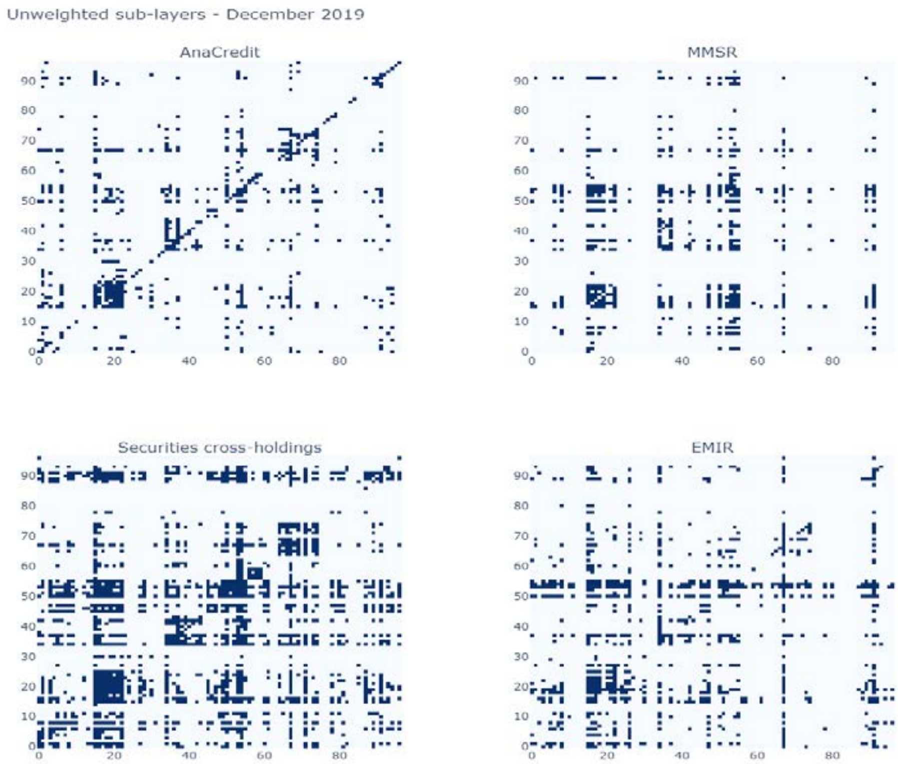


Figure 3

Figure 3 plots the heatmap of the unweighted sub-layers as of December 2019. For confidentiality reasons, the names of the institutions are not displayed, but the nodes have been ordered by country. The differences in the four networks clearly stands out just from a simple visual inspection. It appears that the security cross-holdings layer and the derivative layer are the densest. In the AnaCredit layer, instead, there is a clear pattern of intra-group loans (on the diagonal). Given the way the nodes are ordered, the rectangles we can see around the diagonals represent intra-country clusters, those not on the diagonals are inter-country clusters. One advantage of using heatmaps to visualise a network is that changing the order of the nodes allows to discover and checks for new patterns. For example (not shown here), ordering by total asset size can allow to see whether there are clusters based on size. We could also look at the same picture for the weighted layers, shown in Figure 4. The colour scale here is reversed compared to Figure 3 in order to visualise the low-weight connections: it is easier to see them on a darker background. Compared to the unweighted counterparties, there are only few relevant connections in the layers, highlighted by the lighter colours in the heatmaps.

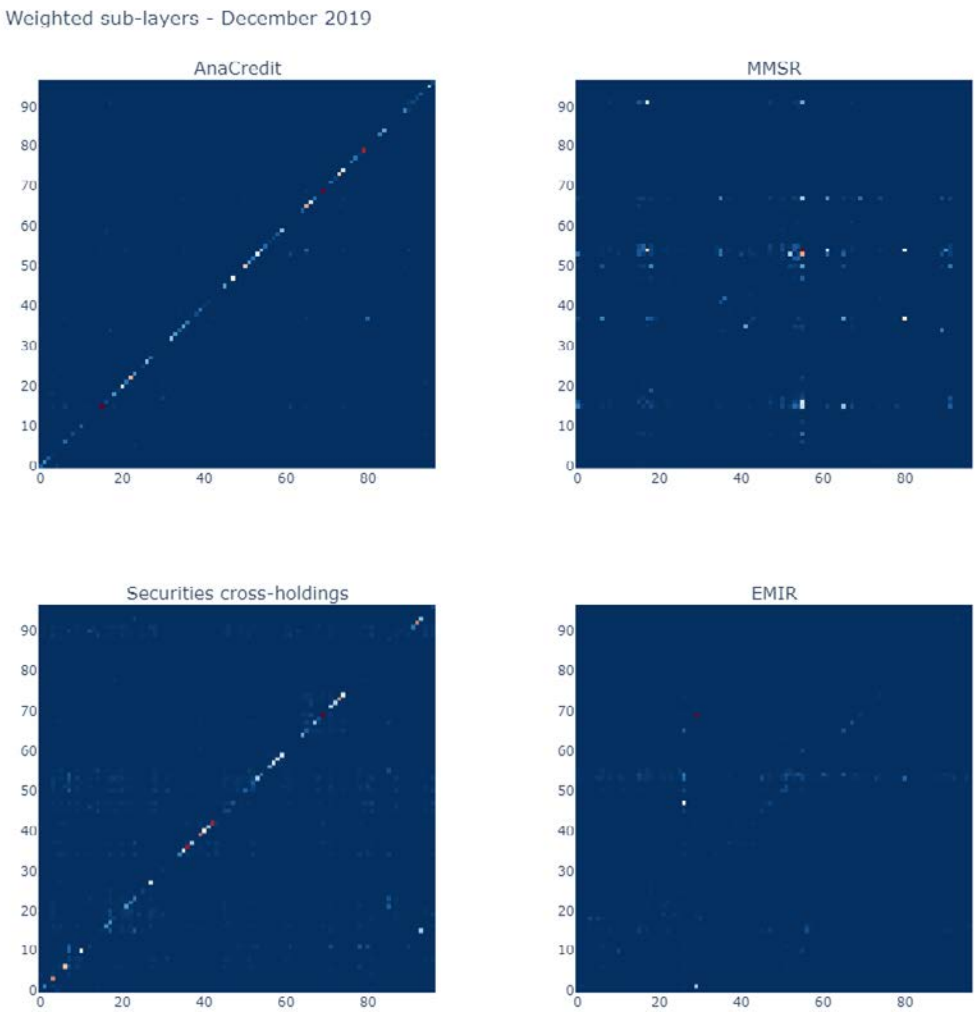


Figure 4

Apart from heatmaps, the tool allows to visualise the network layers in a more standard way. Graphs are usually represented as node-link diagrams, with dots as nodes and lines as the edges among them. The users can use the visualization module to plot the network and see it change over time. Figure 5 has a snapshot of this.

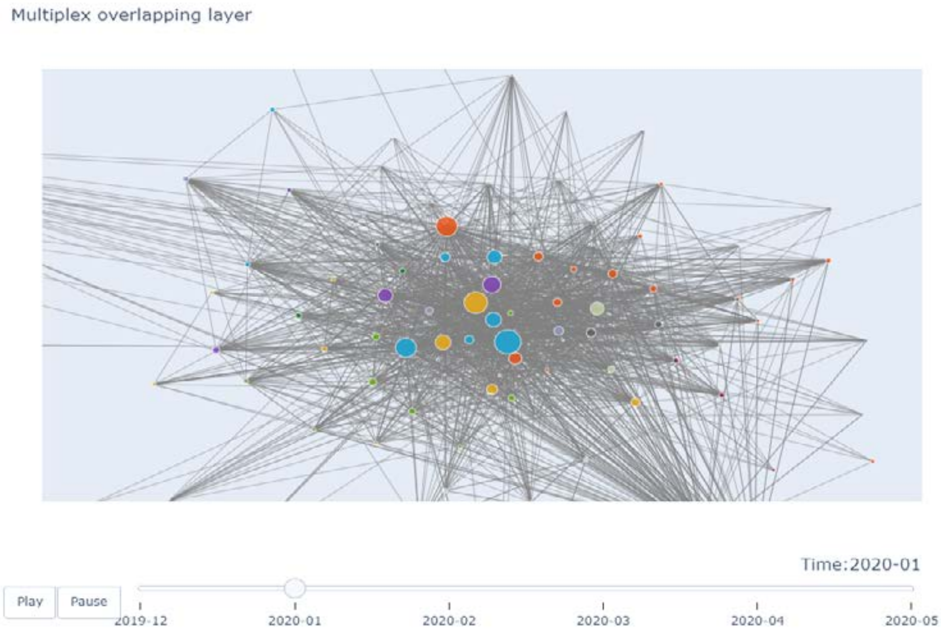


Figure 5

The overlapping layer pictured in Figure 5 is rather dense and, especially in the centre, it is not easy to understand which edge goes where. As stated in Holten and Van Wijk (2009), although node-link diagrams provide an intuitive way to represent graphs, visual clutter quickly becomes a problem in case of graphs comprised of a large number of nodes and edges. One of the most recent trends in network visualisation to address this issue is *edge-bundling*. What edge-bundling does is effectively grouping together edges that go in the same direction. In this way, it is easier to see how the nodes are connected to each other, even when the network is very dense. Figure 6 shows the same network with edge-bundling[11] applied to it. Compared to Figure 5, the structure of the network is now more evident, with nodes in the core and others in the periphery.
Visualising a network is only the first step in its analysis; in order to properly understand its characteristic and dynamics, we need to look at its topology.

[11] The specific bundling algorithm is *hammer_bundle*, a variant of Hurter, Ersoy and Telea (2012) Kernel Density Estimation (KDE) based edge-bundling algorithm. It is implemented through the Python library Datashader. See here for more details on Datashader and here for an intuition on KDE edge-bundling.

Multiplex overlapping layer

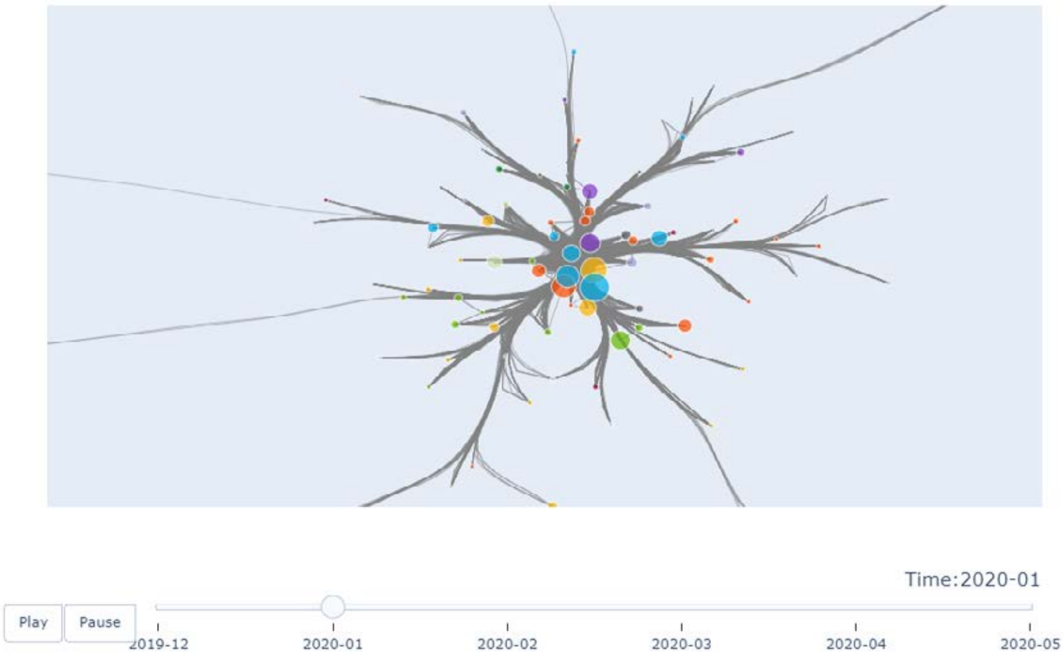


Figure 6

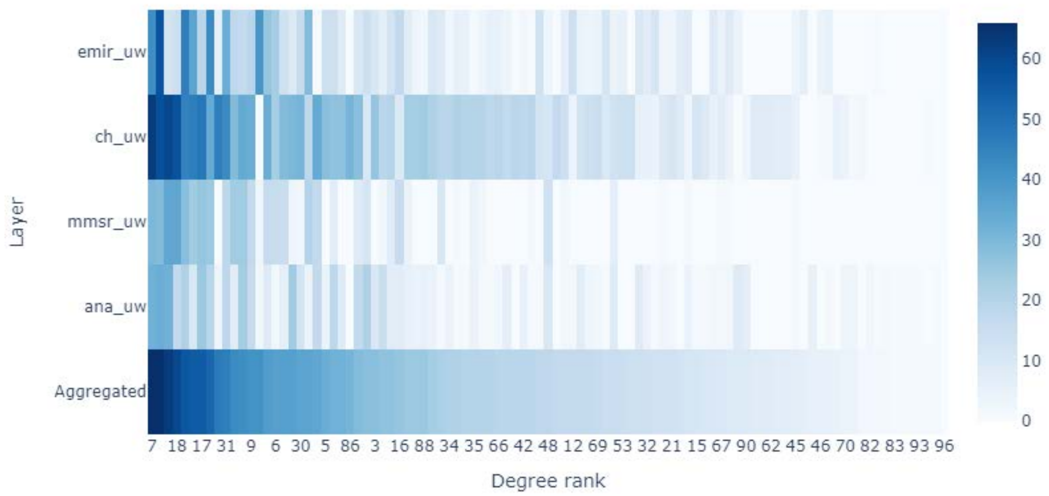
5.2.4. The multiplex topology

One of the first metrics to look at when analysing the topology of a network is the *degree* of a node. The degree tells how many edges are connected to a specific node; in particular, for directed graph (as the one under analysis) we can talk about *in-degree* (edges that come from other nodes) and *out-degree* (edges that go towards other nodes). The degree is a simple and intuitive centrality measure: a node with high degree (either in or out) is central for the network; therefore, this metric can be used to immediately identify the key nodes.

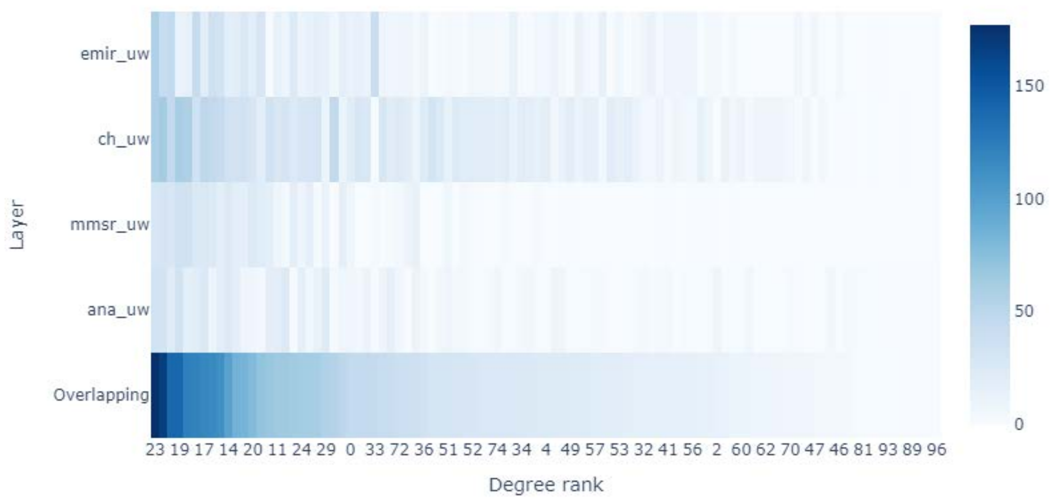
However, since we are analysing a multi-layer graph, the traditional concept of degree centrality needs to be re-framed. When can be a node defined as central? When is it central in all the layers? Or is being central in one layer enough? There is no straight answer; what matters is to find an analytical approach that can give the most complete picture.

Looking at the relationship between the degree distribution in the multiplex layers and the ones in each of the four sub-layers is a first step in figuring out the centrality profile of the nodes. One might wonder whether nodes with high degree in the multiplex are also high-degree nodes in the sub-layers and whether this behaviour changes when we consider the weighted or the unweighted networks. Figure 7 shows the out-degree distributions in the different layers as of December 2019, ordered by the degree ranking of the multiplex. Panel a) focuses on the unweighted layers. Looking at the different distributions, it appears that the cross-holdings layer (ch_uw) is the one resembling the aggregated multiplex layer the most. In other words, high degree nodes in the aggregated layer tend to remain as such in the cross-holding layer. Another feature emerging from a visual inspection is that the out-degree distributions of the loan exposures and secured money market layers and in part that of the derivative layers have a lower range of values (the colour band is lighter) than the one of the aggregated layers. This means that, considering the same institution, this is likely to have more connections in the cross-holdings security layers than in the others. Since we are looking at the out-degree and given how we defined the source and target nodes, the above fact implies that an institution tends to have more exposures towards other institutions in the security market.

a) Ordered out-degree - December 2019 - unweighted layers and aggregated multiplex



b) Ordered out-degree - December 2019 - unweighted layers and overlapping multiplex



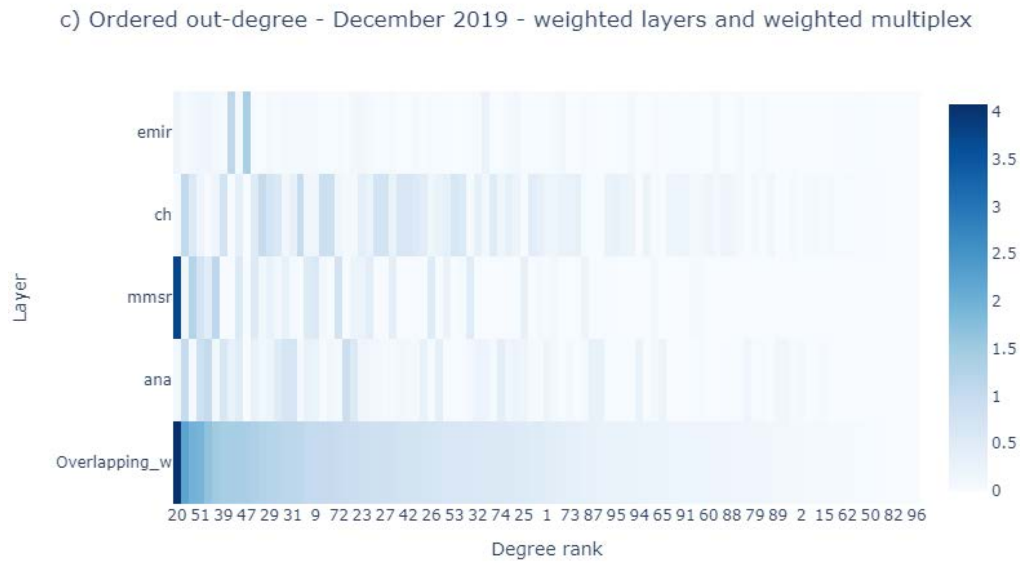


Figure 7

Panel b) confirms the results of panel a) and does not offer new insights. Looking at the weighted layers[12] in panel c) instead, we can spot some differences. First of all, in the weighted overlapping layer we can see a sharper contrast in the colour gradient, meaning there are fewer high degree nodes compared to its unweighted counterparties. Secondly, the secured money market layer distribution appears to be more similar to the multiplex one, compared to its unweighted counterparty. In other words, in terms of absolute number of out connections, the money market layer might appear less correlated with the multiplex layer, but when we look at the size of the transaction, the picture changes. In this regard, it might be useful to look at Figure 8. The two heatmaps show the Kendal correlation[13] as of December 2019 between the different degree distributions.

[12] In a weighted network, the degree is not just the number of edges, but the sum of the weights of the edges.

[13] The Kendall Correlation is a measure of rank correlation: the similarity of the orderings of the data when ranked by each of the quantities. It is a non-parametric alternative to Pearson's correlation (parametric).

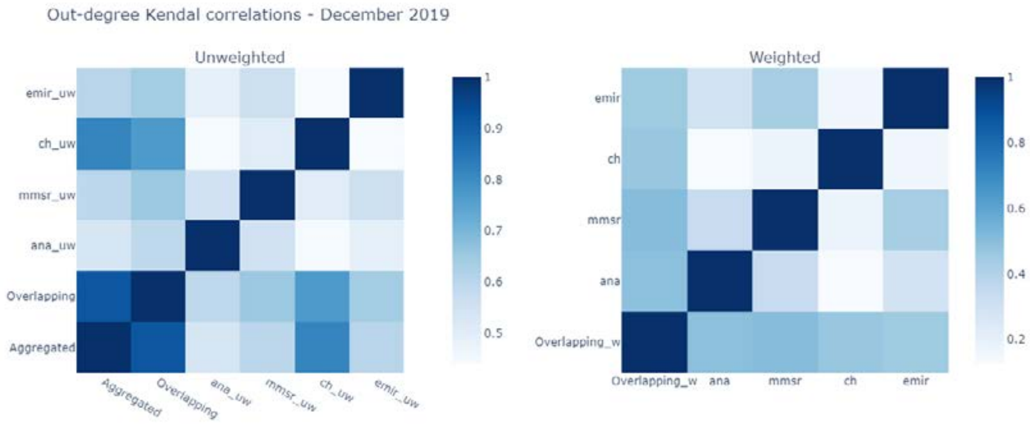


Figure 8

Continuing on the secured money market layer, we can see how, in the unweighted heatmap, mmsr_uw out-degree distribution is not the one with the highest correlation to the degree distributions of the aggregated and overlapping layers: this, instead, is true for its weighted counterpart. Another important difference between the two heatmaps is how the correlation between the multiplex layer and each sub-layer is more diversified in the unweighted networks, compared to the weighted counterparts. This might hint at an inverse correlation between number of transactions in layer and their weight.

Looking at the degree distributions on their own, however, does not tell us clearly whether a high-degree institution in the multiplex layers is so because it equally participates in all sub-layers or instead is a big hub in only one layer. Among the possible metrics to assess this, there is one called the multiplex participation coefficient, defined as:

$$P_i = \frac{M}{(M-1)}\left[1 - \sum_{\alpha=1}^{M}\left(\frac{k_i^{[\alpha]}}{o_i}\right)^2\right]$$

where:

- M is the number of layers.
- $k_i^{[\alpha]}$ is the degree of node i in layer α.
- o_i is the degree of node i in the overlapping layer (also known as overlapping degree).

This coefficient measures whether the links of node i are equally distributed among the M layers or are instead mainly concentrated in just one or a few layers. P_i can have values in the interval [0,1]; in particular, when it is equal to 0 all the edges of i lie in one layer, while $P_i = 1$ only when node i has exactly the same number of edges on each of the M layers. Figure 9 shows the coefficient distribution as of December 2019, for the unweighted (O) and weighted multiplex layers (O^w).

Figure 9

While the unweighted distribution almost looks uniform (if not for a gap between 0.1 and 0.5), the weighted distribution appears to be bimodal. Moreover, while in the unweighted distribution there are over twenty institutions equally participating across all four sub-layers, in the weighted one these institutions decrease to less than five. This means that there are only few institutions with high exposures in all four layers. Another aspect worth understanding is the relationship between the participation coefficient and the degree of a certain node. Figure 10 plots the z-score of the overlapping out degree and the weighted overlapping out-degree over the corresponding multiplex participation coefficients.

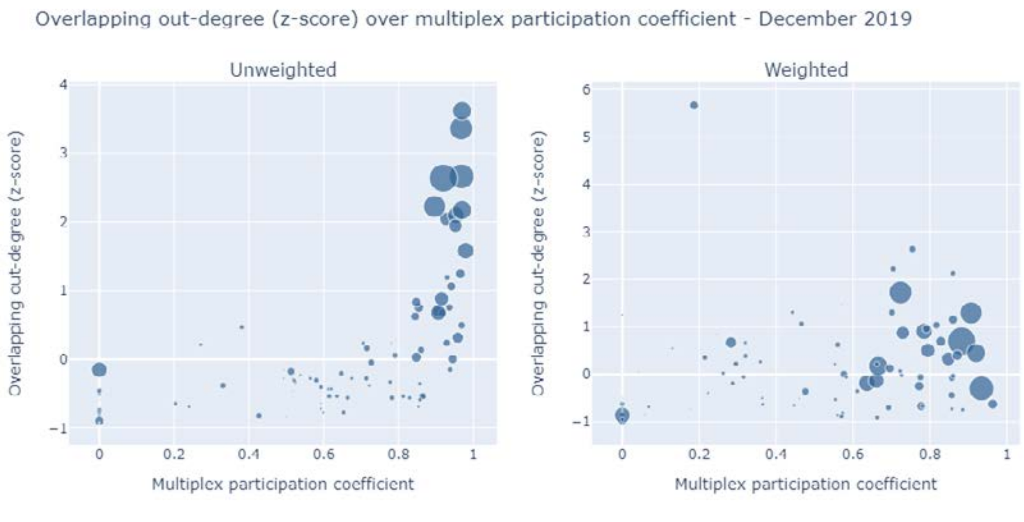


Figure 10 – bubble size represents total assets.

The two scatterplots paint a rather interesting picture. In the unweighted networks, there is a clear positive non-linear correlation between the z-scores and the multiplex participation coefficient. In other words, the more layers an institution has exposures in, the more central it is in the multi-layer network. Moreover, as shown by the size of the bubbles that is increasing moving towards the top-right corner of the graph, there seems to be a positive correlation between z-scores and total assets, as well as multiplex participation coefficients and total assets. In the weighted networks, instead, it appears that the z-scores are almost uniformly distributed across the different values of the participation coefficient. This means that, from an exposures' perspective, the institutions acting in all or only in few markets are not that different in terms of centrality to the network. From the total assets' perspective, there is still a positive correlation between participation coefficients and total assets, while the one with z-scores is not there anymore.

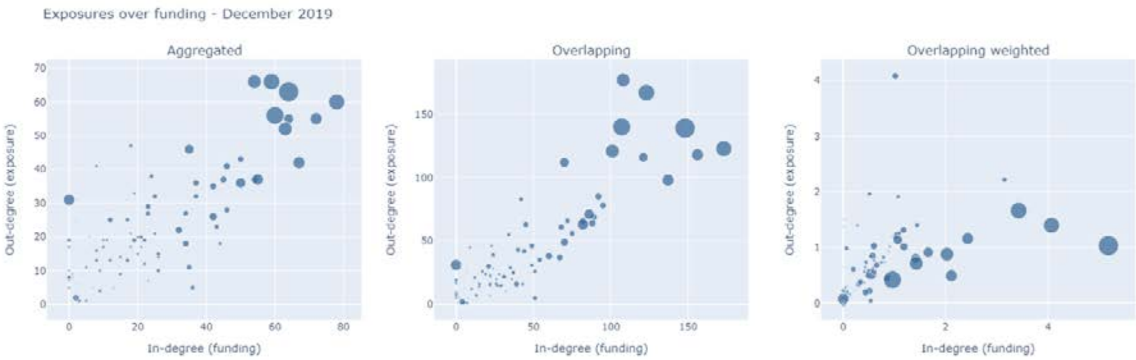


Figure 11 – bubble size represents total assets.

Up to now, we have been looking at the out-degree, thus analysing the network from the exposure side; however, we could also look from the funding perspective with the in-degree. Taking inspiration from Huser and Kok (2019), an interesting relationship to check is the one between exposures and funding, looking at whether institutions with high exposures both in terms of numbers and amount also are the ones receiving more funds, and how this relates to the total assets. Figure 11 provides an interesting insight on the matter. In the aggregated and overlapping multiplex layer there is a positive correlation between funding and exposure: institutions highly connected on the exposure side tend to be so also on the funding side. Moreover, total assets are positively correlated with out and in degree. The overlapping weighted layer tell us a different story: the relationship between exposures and funding is non-monotonic, increasing at first and then decreasing. To better understand this finding, it might be useful to look at each single weighted sub-layer, shown in Figure 12.

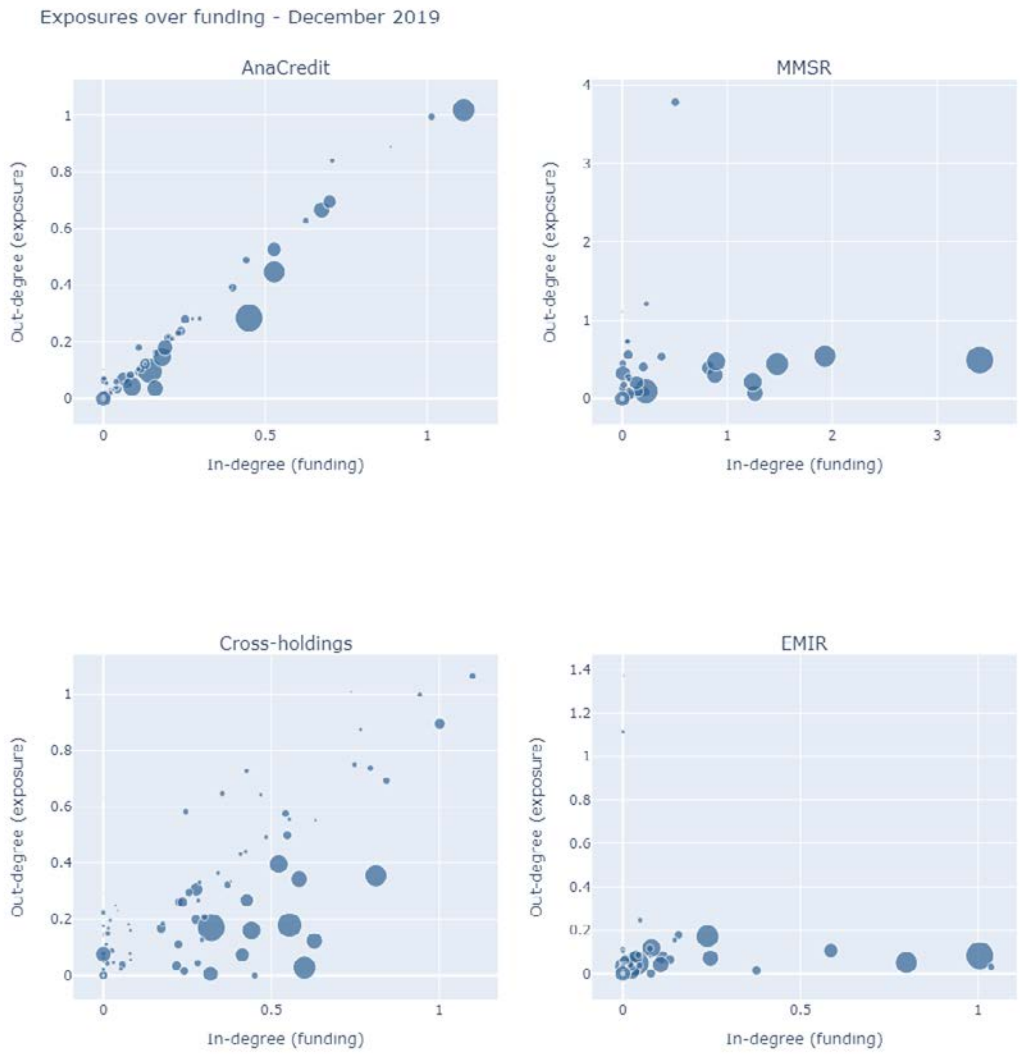


Figure 12 – bubble size represents total assets

While in the AnaCredit and in the security cross-holdings layers there is a positive relationship between funding and exposure, the secured money market layer and the derivative layer do not show such relationship. Instead, it seems that most of the funds received from the network are not redistributed back to it, but likely directed towards other markets. In particular, we can observe nodes with a higher in-degree, compared to the out-degree. One consideration about the cross-holding layer: it appears that the largest (in total asset size) institutions tend act more as issuers of securities, rather than holders. This could be explained by the fact that large institutions issue more securities compared to smaller ones or maybe securities issued by large banks can be more desirable, because less risky, rather than the ones issued by smaller institutions.

So far, we have shown some topological characteristics of the multiplex layers, and how the sub-layers influence them. As a next step, we look at the influence each sub-layer has over the others. The overlapping degree distributions and the multiplex participation coefficients can give an idea of the existence of inter-layers correlations; however, they are not able to disentangle the effect that each layer has over the others. A metric that can give more insights in this regard is the conditional probability of finding an edge e_{ij} in layer α' given the existence of the same edge in layer α:

$$P\left(e_{ij}^{[\alpha']}|e_{ij}^{[\alpha]}\right) = \frac{\sum_{ij} e_{ij}^{[\alpha']} e_{ij}^{[\alpha]}}{\sum_{ij} e_{ij}^{[\alpha]}}$$

For the weighted layers, the probability is:

$$P_w\left(e_{ij}^{[\alpha']}|w_{ij}^{[\alpha]}\right) = \frac{\sum_{ij} e_{ij}^{[\alpha']} w_{ij}^{[\alpha]}}{\sum_{ij} w_{ij}^{[\alpha]}}$$

Figure 13 shows an overview of the different conditional probabilities for the unweighted and weighted version of the four sub-layers.

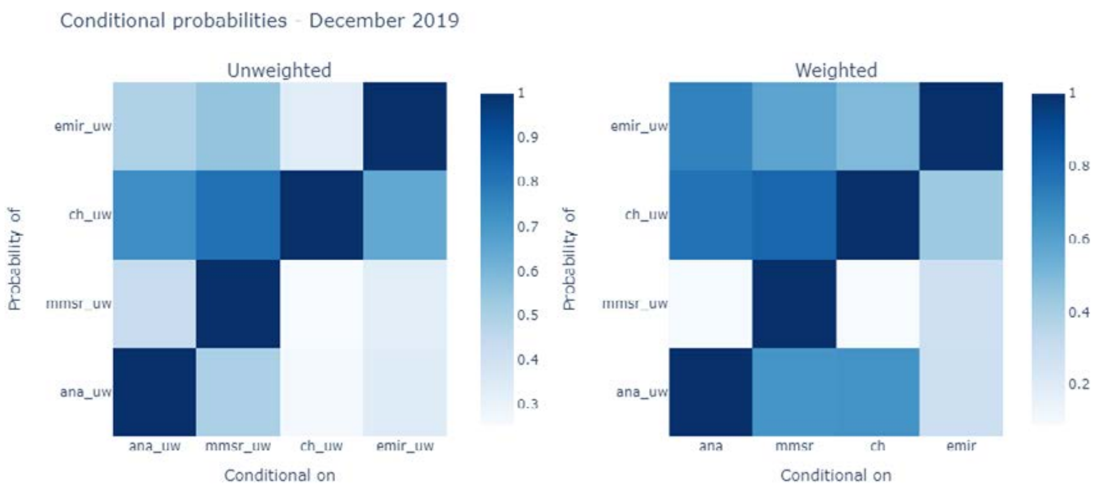


Figure 13

As we can clearly see from the two heatmaps, the probabilities are not symmetric. For example, in the unweighted heatmap, we can see that the probability of an edge being in the cross-holdings layer (ch_uw) conditional on existing in the AnaCredit layer (ana_uw), is much higher than the reverse. Same goes for the probabilities of ch_uw conditional on mmsr_uw. This means that if an institution has a relationship with another on the securities layer, it is very likely that it will also have a relationship on the loan and secured money market layers. However, the inverse is not true. Indeed, this could be explained by the presence of banking groups which are not investment heavy (and thus do not engage on the security layer), but rather focus on more "traditional" banking activity. Overall, the security cross-holdings layer is the one with the highest conditional probabilities, followed by the derivative layer; while AnaCredit and MMSR are the layers doing most of the "conditioning". This is in line with the idea of traditional banking business. Looking at the weighted conditional probabilities, though, we can observe some differences. The secured money market layer is much less influenced by the loan exposure and cross-holdings layers than in the unweighted scenario, while the AnaCredit layer becomes more influenced. Overall, the weight of the relationship affects the conditional probabilities.

6.5. Time dynamics analysis

Up to now, we have been looking at one specific snapshot of our network; however, the network is dynamic, hence it changes over time. Therefore, it is important to have

a framework to observe and assess such changes, as they might contain insights that would be missed otherwise.
One rather straightforward approach to do so is to repeat the above analysis in different time periods. As an example, Figure 14 plots the Kendall correlations of the out-degree distribution (the one in Figure 6) as of each month in our sample.

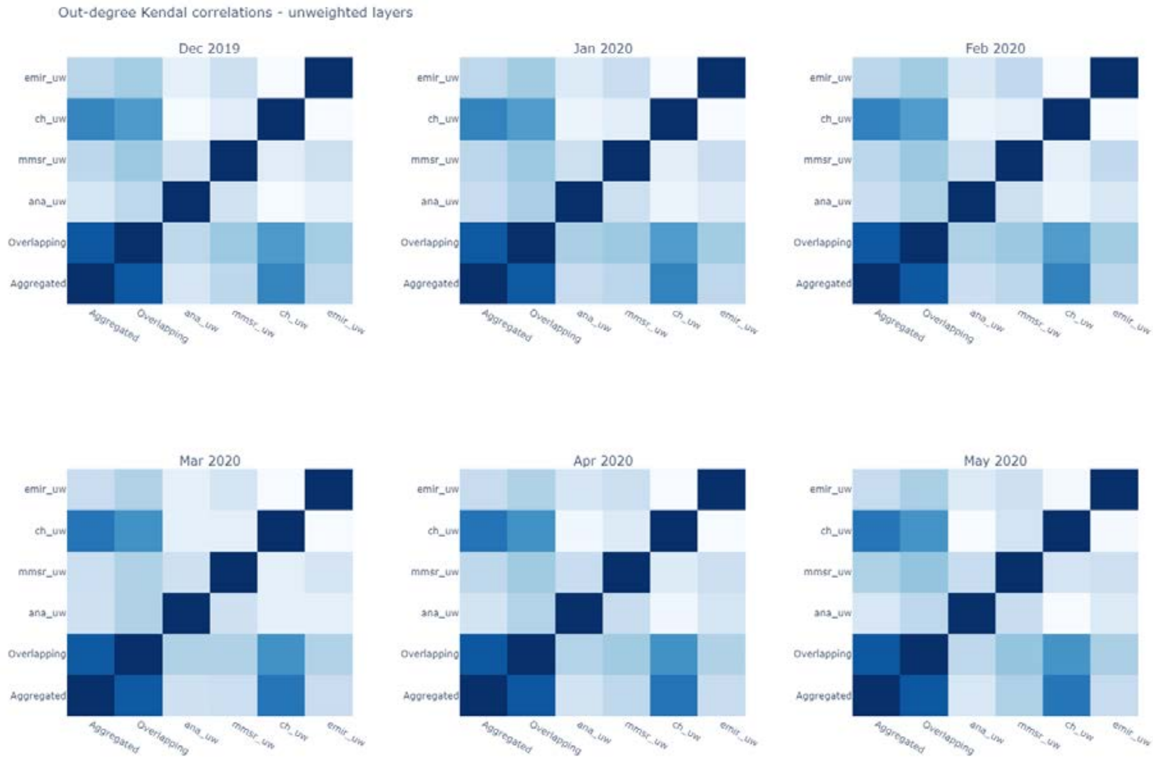


Figure 14

The visualization can be a useful tool to spot relevant differences among time periods. However, looking at this figure in particular, there seem to be none.
A different and maybe more interesting approach is proposed by Goezt and Han (2019). They argue that the majority of measures usually employed to assess network change over time are scalar measures. For example, one can show how the degree of a node changes as time goes by. Instead, they propose to use the cosine similarity to capture not just the change in the node degrees, but its relationship to other nodes. These are vector (or matrix)-based comparisons, rather than scalars, and the authors refer to them as "rewiring" coefficients. With NATkit, it is possible to look at the rewiring coefficients of the multi-layer network over time (Figure 15 and Figure 16).
The absolute rewiring coefficient is simply the cosine distance between the same network (represented in a stacked vector form) in two different time periods. The formula is the following:

$$rc^A(N_{t+1}, N_t) = 1 - \cos\theta_{N_{t+1},N_t}$$

where N is the network under consideration and $\cos\theta_{N_{t+1},N_t}$ is the cosine similarity between the two snapshots of the network. The closer to 1 is the coefficient, the bigger the rewiring in the network; conversely, a coefficient very close to 0 means little distance between the networks. The advantage of this approach to study network rewiring over time is that, being a vector comparison, it allows to compare the network in its entirety, without being forced to choose the perspective of a specific node or node-related metric.

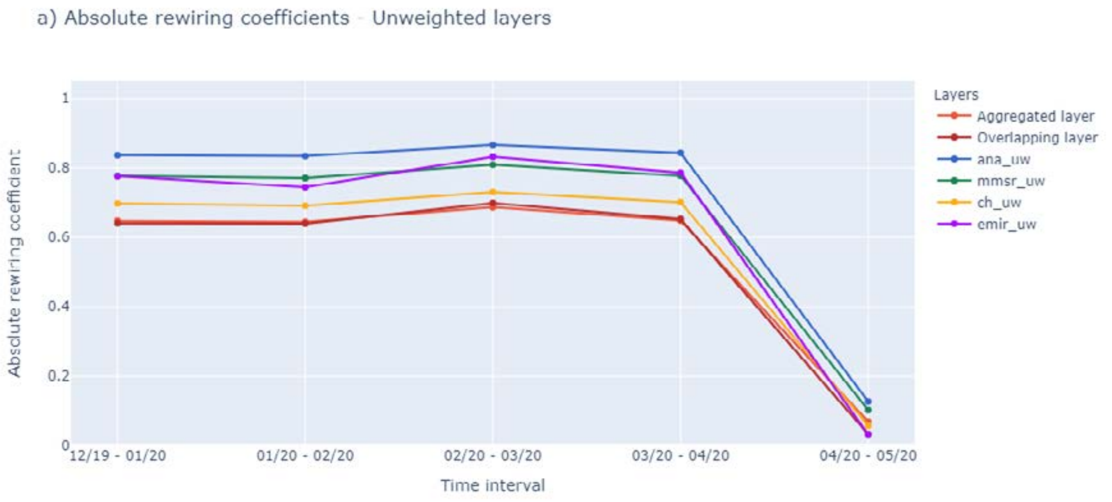


Figure 15

The two plots in Figure 15 show the absolute rewiring coefficients for the different layers in the network. Both unweighted and weighted layers follow the same pattern, with high rewiring observed between December 2019 and April 2020, followed by a sharp drop between April 2020 and May 2020. In order to explain this behaviour, it can be useful to expand the time series of our sample.

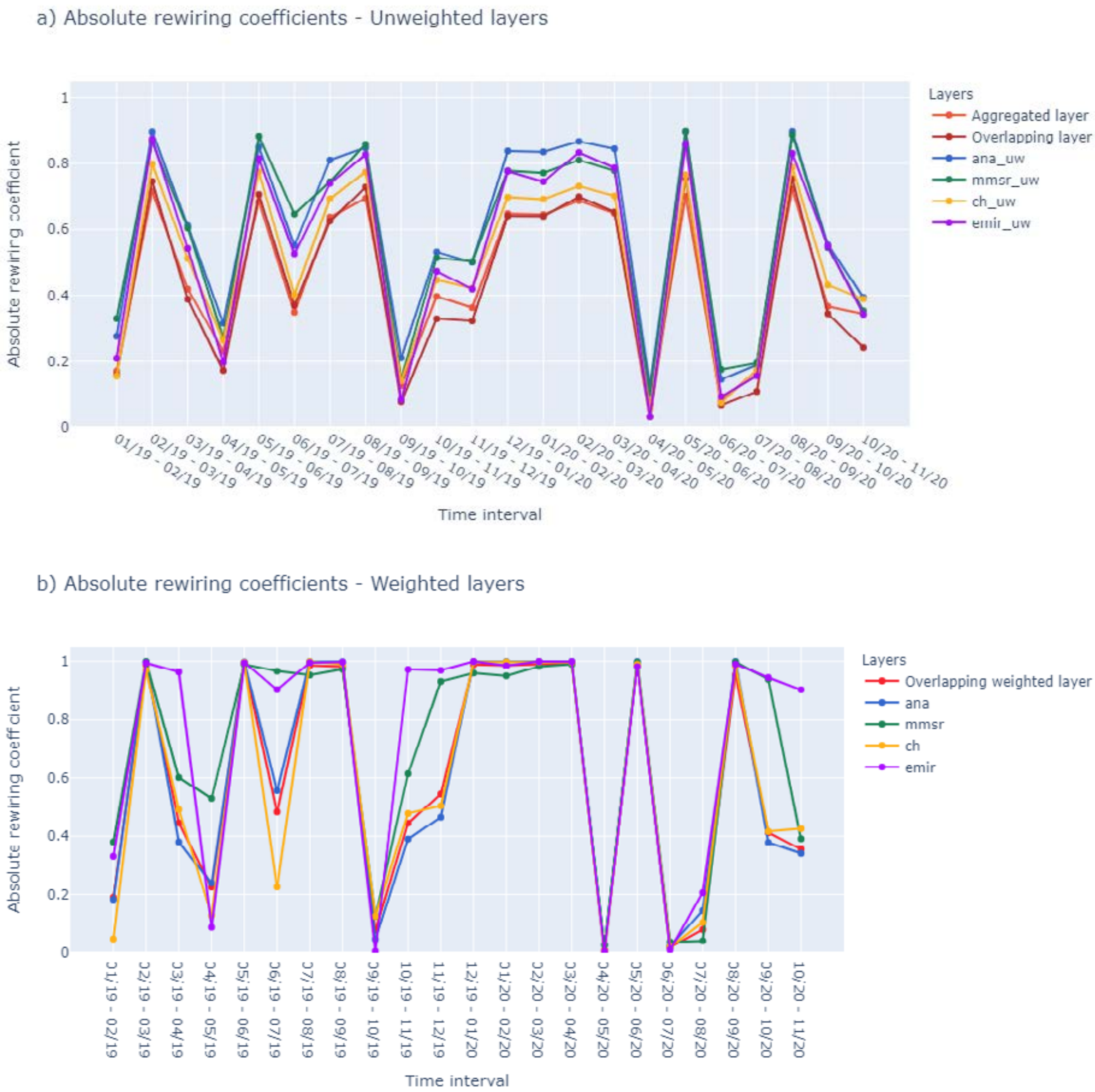


Figure 16

Figure 16 shows the absolute rewiring coefficient over a two-year period. Both the unweighted and weighted time-series show signs of seasonality. Spikes can be observed close to quarter ends: maybe this is due to operations for balance sheet adjustment in view of the reporting obligations. We could check for this hypothesis by simply decomposing the time-series into trend and seasonality[14], using an additive model. Figure 17 shows the results of the decomposition for the aggregated layer.
Looking at the seasonality, we can observe a positive spike in the coefficient for the last month of each quarter. This is in line with our hypothesis of rewiring becoming more intense in view of the reporting obligations. More checks should be performed to validate our guess, but this is out the scope of this work. Regardless of the specific economic interpretation of the patterns of the rewiring coefficient, this metric manages to capture something different from metrics such as network density or average degree, which only show an incomplete picture.

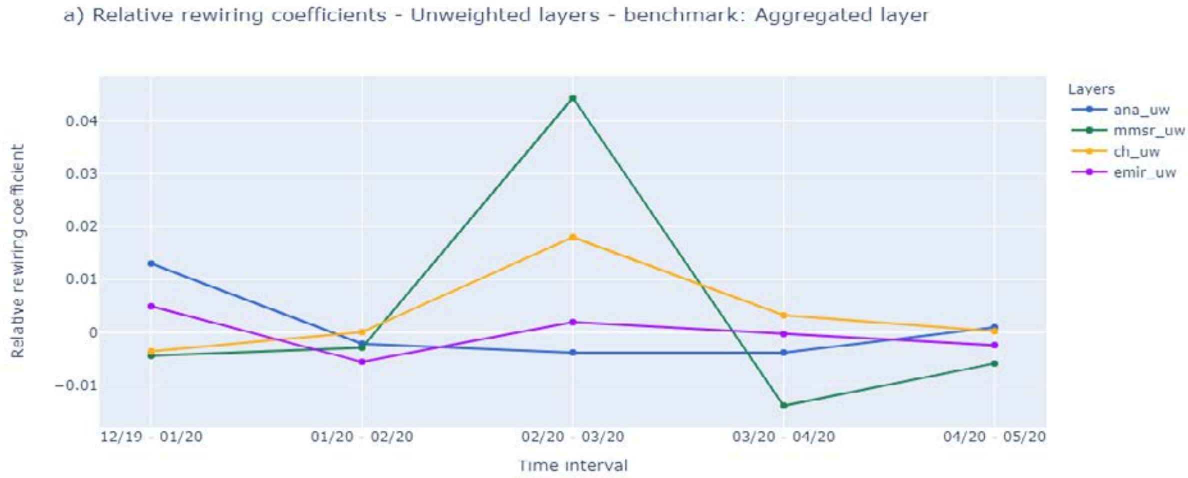
[14] The Python library *statsmodel* offers an easy way to do so.

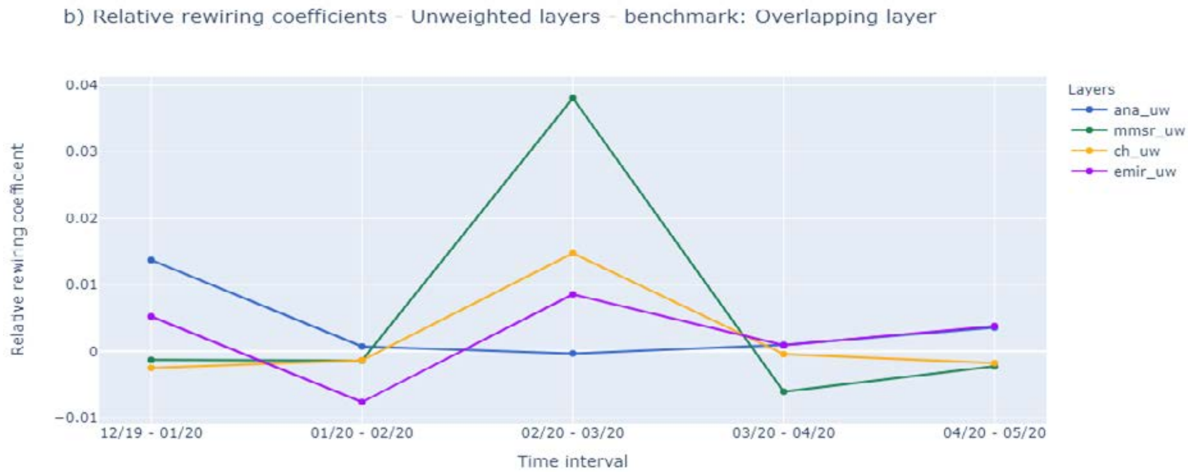
Figure 17

Albeit useful, the absolute rewiring coefficient does not convey the direction of the change. From a vector perspective, the network at $t + 1$ might lie above or below its counterparty at t, pointing in the same or the opposite distance. To identify the direction of change, Goetz and Han (2019) suggest comparing the network change to the equivalent, same-period change in some reference or benchmark network. They call this metrics the relative rewiring coefficient:

$$rc^R(N_{t+1}, N_t, N_{t+1}^B, N_t^B) = \cos\theta_{N_{t+1},N_{t+1}^B} - \cos\theta_{N_t,N_t^B}$$

where N^B is the benchmark network. The relative rewiring coefficient can have values between -1 and 1, where the closer to 1, the more similar is the network to the benchmark one; the closer to -1, the more dissimilar from the benchmark.
Going back to our six-month window, we saw in Figure 15 that each sub-layer and the multiplex layers show very similar rewiring pattern, with higher values of the absolute rewiring coefficient until April 2020, followed by a sharp decline in May. We can now look at the relative rewiring coefficients of each sub-layers, taking as the benchmark the different multiplex representations (aggregated, overlapping and overlapping weighted).





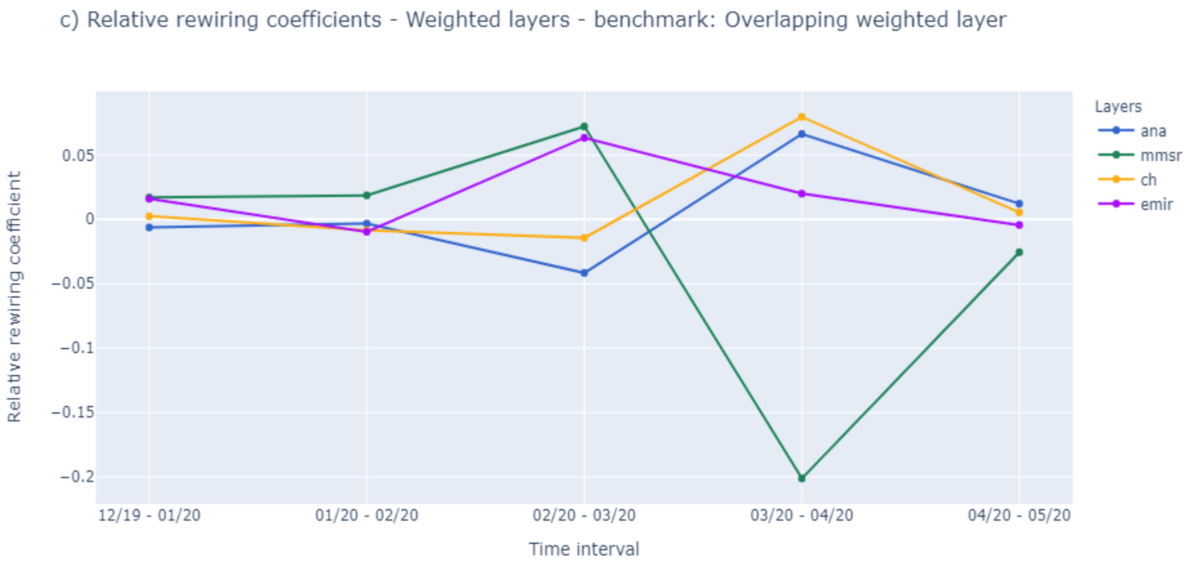


Figure 18

Figure 18 shows the relative rewiring coefficients of the four sub-layers with a different multiplex layer as the benchmark. Panel a) and b) paint a rather similar picture for the unweighted layers. Despite the range of the coefficient being very small and close to 0, we can observe an interesting spike in all the layers, except AnaCredit, between February and March 2020. In that period the all the sub-layers have high absolute rewiring coefficients (Figure 15), but from a relative rewiring perspective, the secured money market layer and the security cross-holdings layer (panel a and b) and the derivative one (panel b only) became more similar to the aggregated and overlapping networks. The loan exposure layer, instead, moved from similarity towards "neutrality", not truly increasing or decreasing its distance from the benchmark layers. Things look different for the weighted layers. The money market layer is still the one with the highest coefficient, however, in the opposite direction of its unweighted counterpart. Moreover, there seem to be a grouping in the layers, with AnaCredit and security cross-holdings moving very similarly and the derivative and secured money market layer moving together in opposite directions from the other two.

5.2.6. Identifying non-linearities

As explained in this showcase analysis, in order to highlight the properties of the multi-layer network, the multiplex layers and the sub-layers are to be analysed and compared together. Indeed, just looking at the aggregated and overlapping (unweighted and weighted) layers on their own might not give the full picture of the complexity of the connections. The literature on the topic has highlighted that multi-layer networks present non-linearities, which would be lost by simply collapsing all the layers together. This, in turn, would lead to wrong results and decisions. Unfortunately, identifying these effects is not straightforward. Nevertheless, with the following metrics, I hope to give the readers a glimpse of or an intuition about them and prove, why multi-layer networks require their own approach and set of adapted metrics.

One property of real-world networks is the tendency of nodes to form the so-called *triangles*, that is, simple cycles involving three nodes. In a triangle ijm, node i is connected to node j and node m, and these two are also connected with each other. Another useful concept is the *triad*. One might think of it as an open triangle: in the above example, node i is connected to node j and node m, but these two are not linked. The clustering coefficient, C_i, allows us to quantify how likely it is that two neighbours of node i are also connected among each other, in other words which proportion of triads close into triangles. Each node has its own clustering coefficient, and averaging them give the average clustering coefficient, C. Table 2 shows the values of the average clustering coefficient for the aggregated multiplex layers and the unweighted sub-layers[15] as of December 2019.

Table 1

Layer	Average Clustering Coefficient
Aggregated	0.601
ana_uw	0.029
mmsr_uw	0.033
ch_uw	0.336
emir_uw	0.054

It is interesting how different the coefficients for the sub-layers are compared to the aggregated layer. The latter has a coefficient almost double the one of the security cross-holdings layer (ch_uw). These results do make more sense when thinking about the definition of the clustering coefficient. Since its the ratio of triangles over triads, in the aggregated layers many of the triads that would not close in a specific layer do close instead. Figure 19 and Figure 20 give some more insight in the distributions of the clustering coefficient.

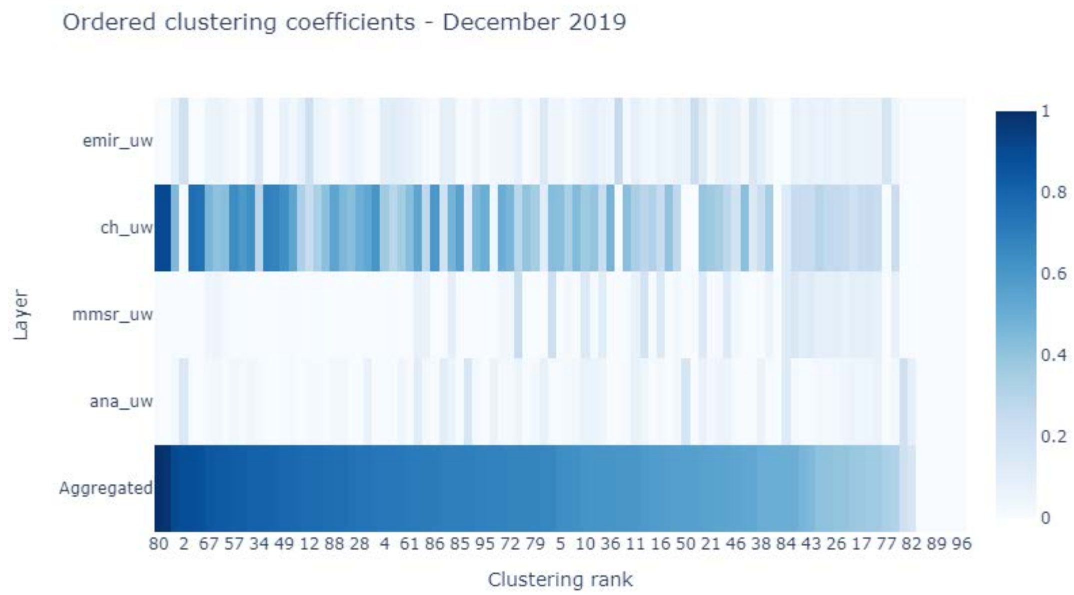


Figure 19

[15] We are focusing on the unweighted layers, because we are interested in the existence of a triangle, rather than its weight.

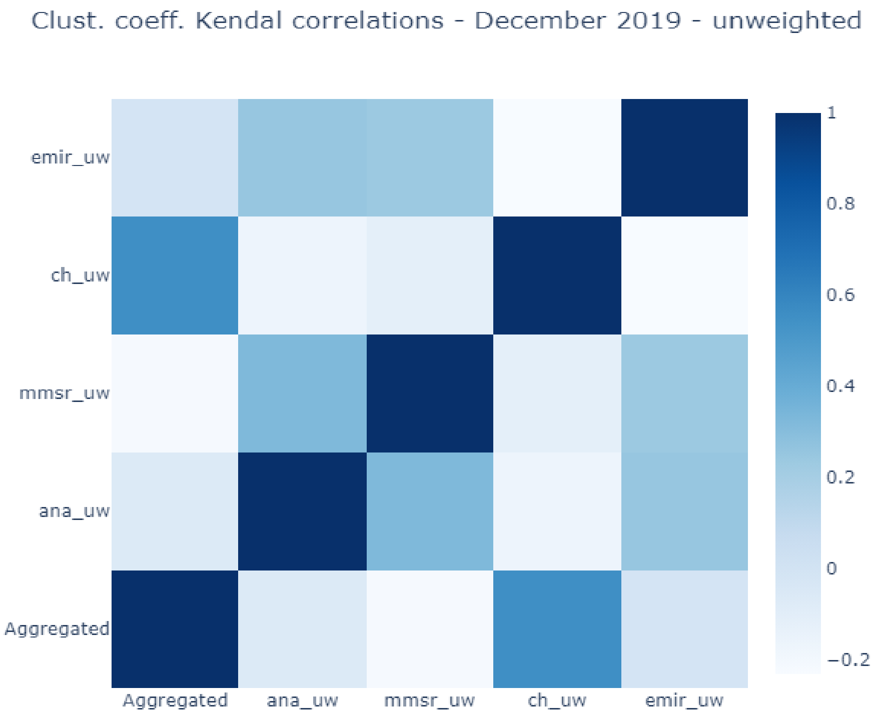


Figure 20

In Figure 19 the contrasts of Table 2 are highlighted. The distribution of the clustering coefficients of the aggregated layer is on a significantly higher range of values, with only the security layer being comparable. Indeed, looking at the Kendall correlations (Figure 20) the aggregated distribution is highly correlated with the one of the security layer.

While these results could be interesting, the traditional clustering coefficient does not consider the multi-layer nature of the network. In terms of computation, the aggregated layer is treated just as any other single layer, and the resulting coefficient tells us very little about the connection across the sub-layers. Therefore, we follow Battiston et al (2014), who computed two modified versions of the clustering coefficient, centred around updated definitions of triangles and triads, more suited for multi-layer systems:

- 2-triangle - a triangle formed by an edge belonging to one layer and two edges belonging to a second layer.
- 3-triangle - a triangle composed by three edges all lying in different layers.
- 2-triad - a triad whose two links belong to two different layers of the systems.

The two new clustering coefficients are defined as:

- C_{1i}: for each node i, the ratio between the number of 2-triangles with a vertex in i and the number of triads centred in i.

- C_{2i}: for each node i, the ratio between the number of 3-triangles with node i as a vertex, and the number of 2-triads centred in i.

The advantage of these two new measures is that they take into account all the sub-layers together, instead of simply looking at the aggregated layer. The new concepts of triangles and triads allow to differentiate between a triangle that is all in one layer and one who is in two or three layers. This is important because the connection between two banks in the loan exposure layer is not the same connection in the derivative layer. Shock propagation can be a practical application where such a difference is important: assuming there is a shock in one specific market, we need to know that it will propagate differently if the next connection is on the same market or on another one.

For the month of December 2019, the average clustering coefficient of the aggregated layer is 0.601. The average C1 and C2 clustering coefficient for the multi-layer unweighted network are 0.136 and 0.244, respectively. This means that simply aggregating the sub-layers together would over-estimate the clustering of the network, because it would overlook the non-linear interactions existing across the layers.

6. Conclusions

After the financial crisis, the ECB accelerated the collection of granular statistical and supervisory datasets. Each of them can bring new insights into the function of the financial markets and offer new ways to assess risks for their stability and thus help policy makers in their decisions. We argue, however, that only looking at these asset classes simultaneously can account for non-linearities that are caused by the interconnectivity of those markets. We therefore constructed a new dataset that integrates these granular data and which we use to build a multi-layer dynamic network which covers loans, securities, derivatives, and money market transactions of the significant banking groups in the euro area.

The newly created network is constructed in a modular and flexible way. This way researchers can use the definition of banking groups they need, or they select only those layers they need for the analysis of the credit market For example, for some use cases a supervisory definition is needed, while others might want to look at pure ownership structures. In different cases, users might want to focus only on the loan and securities layers in their research and disregard the other layers.

No matter how the multi-layer dynamic network is used, it becomes clear that new tools are needed to deal with the interconnections of the different layers. We therefore describe how we put together useful tools in a Python package NATkit, that supports the analysis of the combined layers. As we have shown in this paper, multi-layer dynamic networks need an ad-hoc analytical approach that can properly capture and describe their topology. A multi-layer network is not just the sum of its sub-layers; therefore, its analysis should always encompass both individual layers and the multiplex layers. Each tells a part of the story, and both sets are needed to get the full picture. With NATkit, we tried to establish the framework to enable this, taking from the relevant literature the best suited approaches and metrics for the type of data under consideration.

We hope that the code and the documentation on the construction of the network as well as the toolkit, spark more research in this topic and accelerates the interest in using the granular datasets. The work on these data, however, is far from over. Our work has shown that unique, reliable, and timely master and meta data are key for the construction of such a network. Without them the granular datasets can not be used to their full potential. Therefore, many initiatives on data integration and further development of master data are going on in the background of the ESCB: working groups are continuously strive to enhance the coverage of identifiers in the datasets, which makes it easier to integrate the data. At the same time, several project as The Integrated Reporting Framework (IReF) and the Banks' Integrated Reporting Dictionary (BIRD)[16] will pave the way to achieve true semantic integration for granular banking data in the ESCB, which will greatly enhance the data quality of such projects as the multi-layer network.

[16] See also the ESCB long-term strategy for banks' data reporting.

References

Aldasoro, I., & Alves, I. (2018). Multiplex interbank networks and systemic importance: Anapplication to European data. *Journal of Financial Stability, 35*, 17-37.

Allen, F., & Gale, D. (2000). Financial Contagion. *Journal of Political Economy*.

Bargigli, L., di Iasio, G., Infante, L., Lillo, F., & Pierobon, F. (2015). The multiplex structure of interbank networks. *Quantitative Finance, 15*(4), 673-691.

Battiston, F., Nicosia, V., & Latora, V. (2014, March 12). Structural measures for multiplex networks. *Physical Review E, 89*(3).

Draghi, M. (2019). Macroprudential policy in Europe. *Fourth annual conference of the ESRB.* Frankfurt am Main.

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae* , 290-297.

Farkas, M. (2020). Network Topology of Target2 Interbank Payments. *ECB Graduate Programme Paper*.

Freixas, X., Parigi, B. M., & Rochet, J.-C. (2000, August). Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank. *Journal of Money, Credit and Banking, 32*(3), 611-638.

Han, Y., & Goetz, S. J. (2019, July 24). Measuring network rewiring over time. (I. Sendiña-Nadal, Ed.) *PLoS ONE, 14*(7).

Holten, D., & van Wijk, J. J. (2009). Force-Directed Edge Bundling for Graph Visualization. (H. -C. Hege, I. Hotz, & T. Munzer, Eds.) *Eurographics/ IEEE-VGTC Symposium on Visualization, 28*(3).

Hurter, C., Ersoy, O., & Telea, A. C. (2012, June). Graph bundling by Kernel Density Estimation. *EUROVIS 2012, Eurographics Conference on Visualization*, 865-874.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. (E. Estrada, Ed.) *Journal of Complex Networks, 2*, 203-271.

Kok, C., & Hüser, A.-C. (2019). Mapping bank securities across euro area sectors: comparing funding and exposure networks. *ECB Working Paper Series, 2273*.

Kurant, M., & Thiran, P. (2006, April). Layered Complex Networks. *Physical Review Letters, 96*.

Poledna, S., Molina-Borboa, J. L., Martínez-Jaramillo, S., van der Leij, M., & Thurner, S. (2015, October). The multi-layer network nature of systemic risk and its implications for the costs of financial crises. *Journal of Financial Stability, 20*, 70-81.

Réka, A., & Barabási, A.-L. (2002, January 30). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics, 74*(1).

Spiaggiari, M. (2020). A journey from data management to data analysis with Anacredit. *ECB Graduate Programme Paper*.

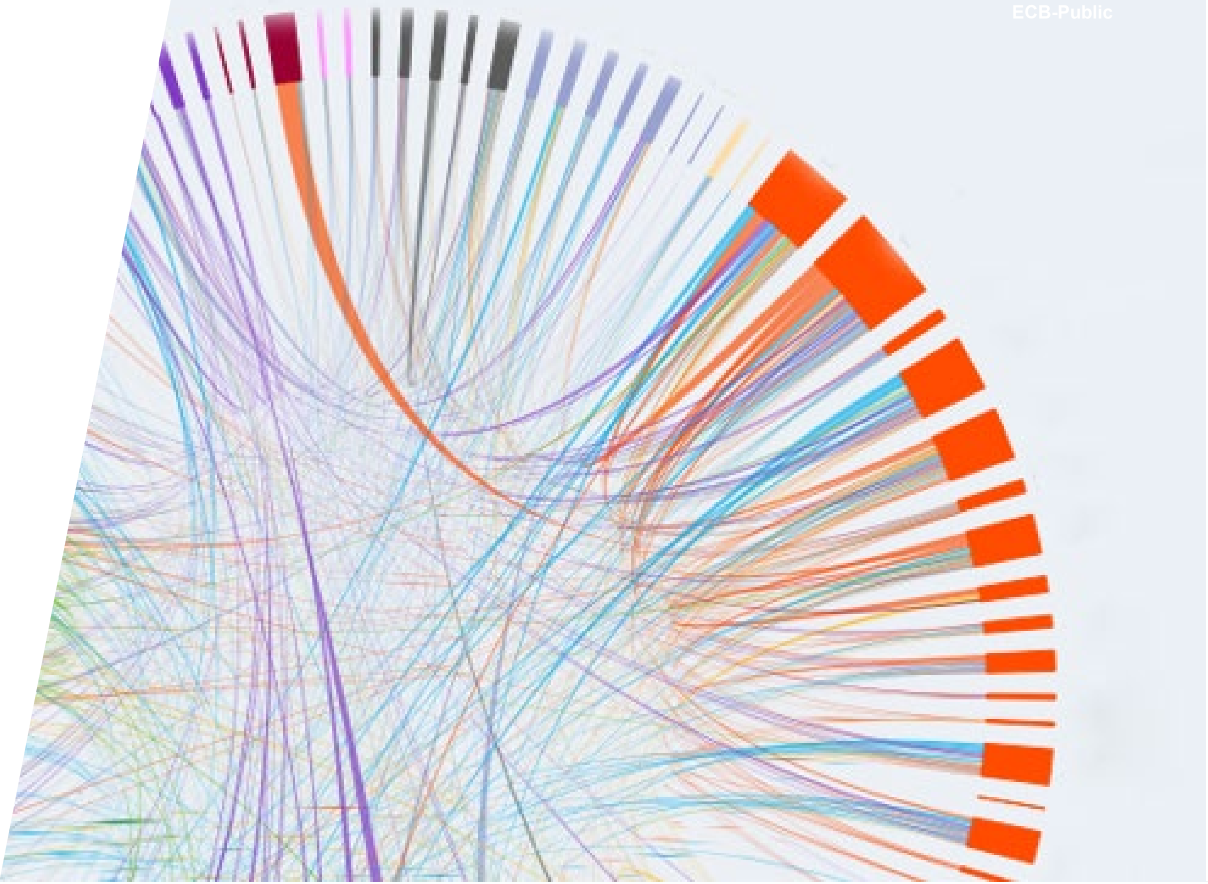
ECB-Public

EUROPEAN CENTRAL BANK

EUROSYSTEM

A multi-layer dynamic network

For significant European banking groups



14/02/2022

Annalaura Ianiro
Directorate General Macroprudential Policy and Financial Stability

Jörg Reddig
Directorate General Statistics

Presentation: A multi-layer network

Granular data at the ECB

The ECB has a rich set of granular data at their disposal

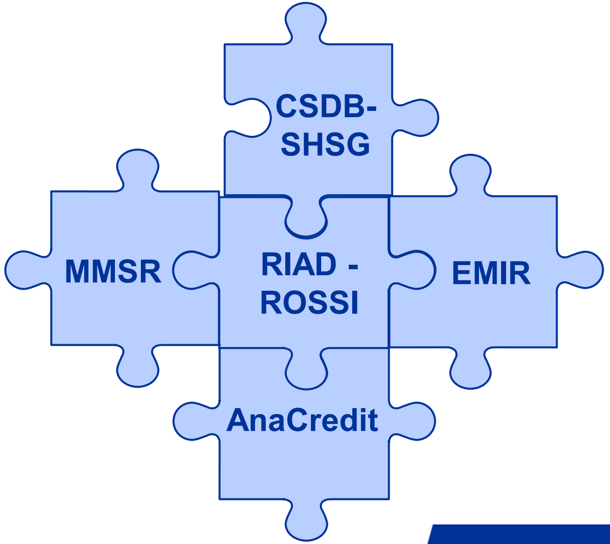
- Data collected by ECB and other European institutions
- Covers around 95% of the asset side of banking groups

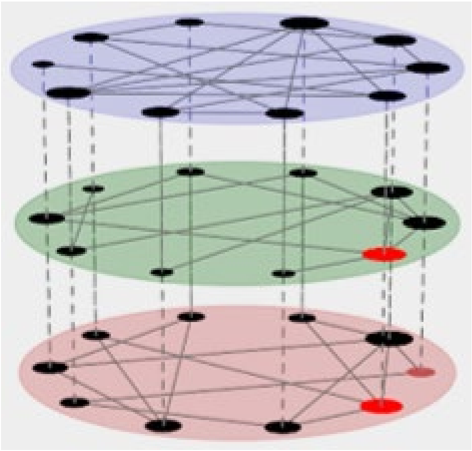
The challenge lies in combining these datasets

- Different codelists and dictionaries used
- Different taxonomies (statistical vs supervisory)
- Need for unique, high-quality set of master data

Granular data at the ECB

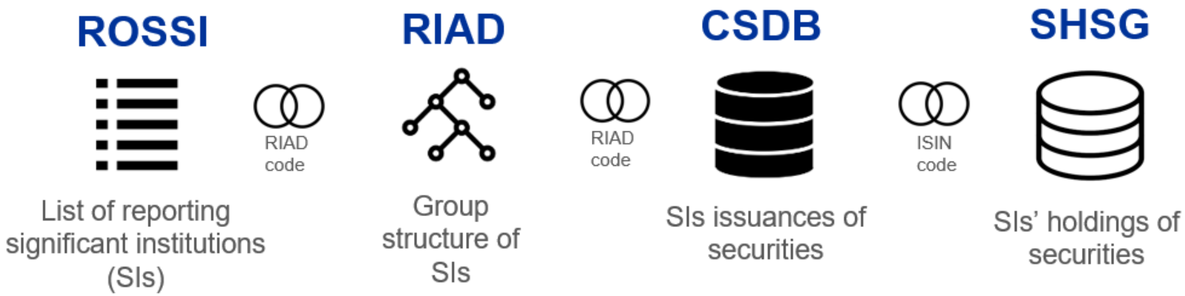
- For the first time it is possible to integrate and analyse the **interconnections of banking groups** in terms of *securities, loans, money market trades* and *derivatives.*



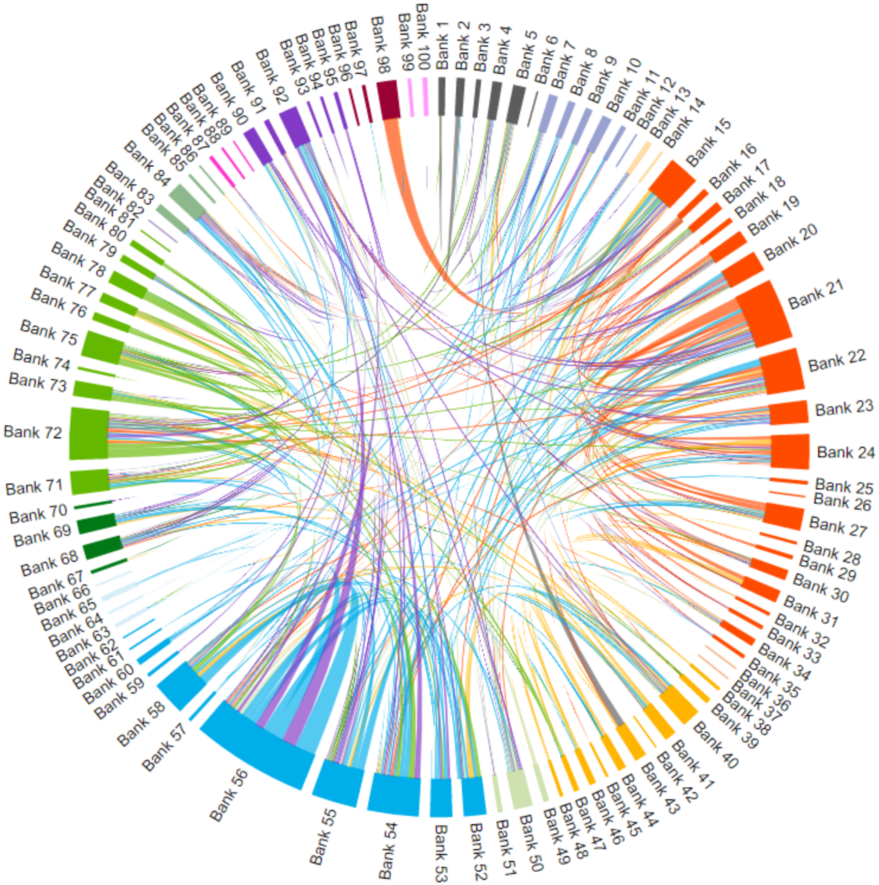


Building a multi-layer network

Integration of data – example of the securities layer



Network



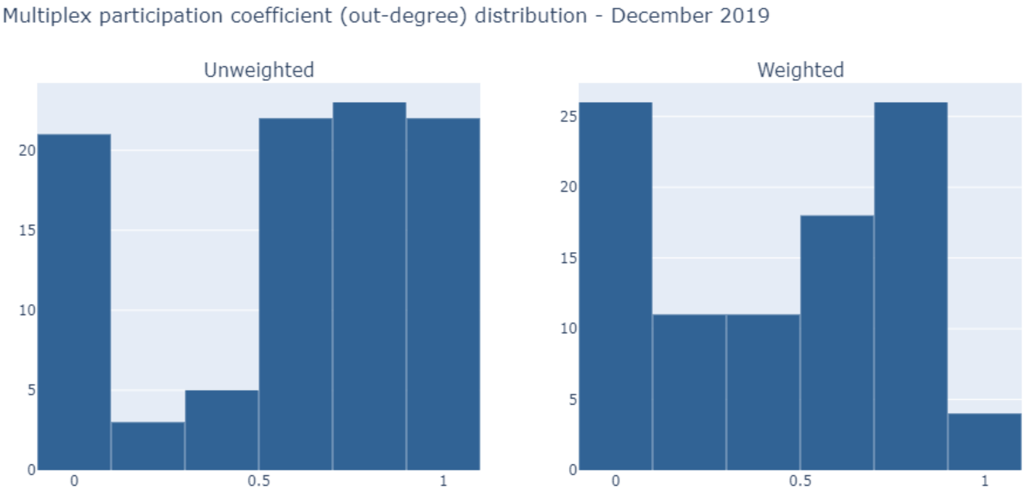
Bank 1
Bank 2
Bank 3
Bank 4
Bank 5
Bank 6
Bank 7
Bank 8
Bank 9
Bank 10
Bank 11
Bank 12
Bank 13
Bank 14
Bank 15
Bank 16
Bank 17
Bank 18
Bank 19
Bank 20
Bank 21
Bank 22
Bank 23
Bank 24
Bank 25
Bank 26
Bank 27
Bank 28
Bank 29
Bank 30
Bank 31
Bank 32
Bank 33
Bank 34
Bank 35
Bank 36
Bank 37
Bank 38
Bank 39
Bank 40
Bank 41
Bank 42
Bank 43
Bank 44
Bank 45
Bank 46
Bank 47
Bank 48
Bank 49
Bank 50
Bank 51
Bank 52
Bank 53
Bank 54
Bank 55
Bank 56
Bank 57
Bank 58
Bank 59
Bank 60
Bank 61
Bank 62
Bank 63
Bank 64
Bank 65
Bank 66
Bank 67
Bank 68
Bank 69
Bank 70
Bank 71
Bank 72
Bank 73
Bank 74
Bank 75
Bank 76
Bank 77
Bank 78
Bank 79
Bank 80
Bank 81
Bank 82
Bank 83
Bank 84
Bank 85
Bank 86
Bank 87
Bank 88
Bank 89
Bank 90
Bank 91
Bank 92
Bank 93
Bank 94
Bank 95
Bank 96
Bank 97
Bank 98
Bank 99
Bank 100

Analysing a multi-layer network

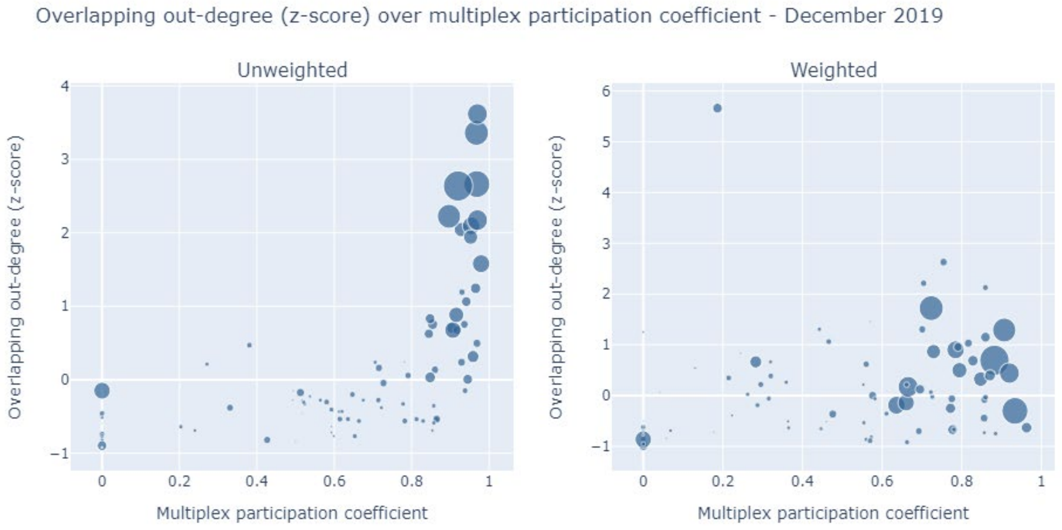
- Network Analytics Toolkit, developed to **analyse multi-layer dynamic networks.**
- Python library with **three modules**:
  - Visualization
  - Topology
  - Shock propagation (early development stage)
- Available as an open source tool within the ECB

Analysing a multi-layer network: multiplex topology

- A high degree node in the multiplex layer can be such because:
  - it **equally participates** in all sub-layers.
  - it is a big hub in **just one layer**.
- The **multiplex participation coefficient** tells if the links of a node are distributed among the sub-layers or are concentrated in just one or few layers.



Analysing a multi-layer network: multiplex topology



- What are the central institutions?
- In the unweighted network, the more layers an institution has exposures in, the more central it is.
- In the weighted network, being active in all or only few markets does not make a big difference in terms of degree centrality.
- Positive correlation between multiplex participation coefficients and total assets.

Conclusion and outlook

- The network opens **new areas of research** not possible before
  - Code and detailed documentation is openly available inside the ECB
  - Can be used as a starting point for data integration projects
- **Important work in the background:**
  - Data quality improvements, harmonising identifiers, building common dictionaries
- Future improvements:
  - Work on **semantic integration** of datasets and **integrated reporting** (IReF project)
  - Adding **new layers** (i.e. SFTDS).

Thank you for your attention!

Contact and Disclaimer

Annalaura Ianiro

Annalaura.Ianiro@ecb.europa.eu

annalauraianiro

Jörg Reddig

Joerg.Reddig@ecb.europa.eu

joergreddig

The views expressed are those of the authors only and do not necessarily reflect those of the European Central Bank.