Master of Science in Big Data & Business Analytics

Executive edition 2015



Analytics applied to finance

*Proposal of a methodology for increased efficiency in the Systemic Risk analysis for the OTC markets.*



Authors: Karim Rezc

Alfonso Campos

# Introduction

Nowadays National Competent Authorities (NCAs) have Access to all the trades reported by the market participants, this is due to the forced application of several regulatory frameworks like MIFID, EMIR, SFTR, REMIT, etc.

In this paper the objective is to propose a methodology for Financial regulators in order to supervise the Over The Counter markets in a more efficient way, and prevent any kind of systemic risk situation like the recent financial crisis that happened during 2008.

Having in mind the massive amount of information that the six European trade repositories receive in a daily basis, all the European NCAs will have to develop big analytics departments in order to perform their supervisory duties based on the data received.

A regulatory ecosystem is composed by a market infrastructure that is known as a Trade Repository (TR), in Europe there are six TRs authorized by the European Securities and Market Authority (ESMA), and at the same time these TRs will centralize all the transactions reported by the market participants in order to provide a single access point to all the European NCAs. During this paper we will discuss only the case of EMIR

The European Markets Infrastructure Regulation (EMIR) was issued in response to G20 Pittsburgh commitments to financial reform and regulates the authorization, registration, organizational requirements and supervision of European Trade Repositories (TRs). EMIR legislation (level 1 text) became law in July 2012.

According to duties described on Article 9 of EMIR:

*“Counterparties and CCPs shall ensure that the details of any derivative contract they have concluded and of any modification or termination of the contract are reported to a trade repository registered in accordance with Article 55 or recognised in accordance with Article 77. The details shall be reported no later than the working day following the conclusion, modification or termination of the contract. EN L 201/20 Official Journal of the European Union 27.7.2012”*

*The adoption of the regulatory and implementing technical standards on OTC derivatives, central counterparties and trade repositories mentioned the following points that are relevant for the business case of this project:*

“**Adoption of the regulatory and implementing technical standards for the Regulation on OTC derivatives, central counterparties and trade repositories – 19.12.2012**

On 19 December 2012, the European Commission has adopted nine regulatory and implemented technical standards to complement the obligations defined under the Regulation on OTC derivatives, central counterparties (CCPs) and trade. They were developed by the European Supervisory Authorities and have been endorsed by the European Commission without modification.

The adoption of these technical standards finalizes requirements for the mandatory clearing and reporting of transactions, in line with the EU’s G20 commitment made in Pittsburgh in September 2009

On 9 February 2012, the European Parliament and the Council reached an important agreement on a Regulation for more stability, transparency and efficiency in derivatives markets. It was a key step in the effort to establish a safer and sounder regulatory framework for European financial markets.

On 4 July 2012, the Regulation on OTC Derivatives, Central Counterparties and Trade Repositories (known as "EMIR" - European Market Infrastructure Regulation) was adopted and entered into force on 16 August 2012. This was a major development which enables the European Union to deliver the G20 commitments on OTC derivatives agreed in Pittsburgh in September 2009.

The Regulation ensures that information on all European derivative transactions will be reported to trade repositories and be accessible to supervisory authorities, including the European Securities and Markets Authority (ESMA), to give policy makers and supervisors a clear overview of what is going on in the markets[[1]](#footnote-1).”

# Scope of the problem to be solved

The challenging part of this new regulatory framework, it is not only the creation of complex and reliable technical infrastructure like Trade Repositories, but for the most part, the regulatory implementation implies the adaptations and infrastructures that NCAs will have to build in order to be capable of analyzing and detect systemic risk, market abuse, etc.

It is worth to note that TRs receive millions of transactions in a daily basis, therefore the processing capacity of an NCA (in fact they are the data consumers) will have to be increased, so potentially the learning curve of the implementation will not be easy. In fact the problems that the implementation of the regulation implies are that not all the European NCAs have the same budget or infrastructure.

For instance the differences can be significant across the different NCAs; Bafin (the German regulator) is one of the most bigger NCAs in Europe, but there are regulators like the Malta Financial Authority that does not have the resources or technical infrastructure that Bafin does.

Having said that, our approach for supervising systemic risk will be to avoid analysis and processing of all the data provided by the TRs (which can be in the hundreds of GBs daily), and instead, a first analysis of the data will give some light about the potential points of systemic risk or market bailout producers on the OTC markets.

# Objective

The scope of this analysis is to propose a process for NCAs that will enhance an efficient supervision for detecting systemic risk. With this process, NCAs will be able to find out the full picture of the systemic risk in the bilateral over the counter trades (from herein by OTC trades), our proposal is based on the hypothesis that systemic risk is concentrated in a small group of market players, therefore the supervisory efforts should be centered on analyzing the transactions reported by these market players.

The assumptions that have to be taken into consideration are the following:

* A systemic risk concentrator will be considered as the entity that executes the major number of trades with different counterparties
* This analysis is not measuring the level of exposure of every derivative contract executed (notional amount), but only the OTC trades
* Systemic risk is considered as a market meltdown produced by the propagation effect due to the default or credit event of a systemic risk concentrator.
* An additional conclusion of this methodology will be that the full picture of the risk concentration will be illustrated, and it would be interesting for NCAs to record these pictures over time in order to understand how the OTC market is behaving.
* The scope of this study is just to propose an improvement to the systemic risk analysis , then the infrastructure details that will have to implement every NCA will not be mentioned.

# Methodology proposed

The methodology proposed will be composed of the following stages:

* **Data extraction.** Nowadays NCAs receive daily reports of the different TRs that include all the information of the different counterparties across Europe, the data extraction process for NCAs will be to receive all the data. This step will not be illustrated in this document, thus the baseline of this process will be that an NCA has already received all the reports from all the different Trade Repositories.
* **Data Transformation.** In this stage, we will assume that an NCA will load all the data into a simple Database, and the NCA will have to run a query in order to select the relevant data for all the reports received. In this stage the NCA will split the data into ETD (exchange traded derivatives) and OTC (over the counter trades), then the NCA will extract from the reports only the counterparties of each trade OTC trade, once this step is achieve then the analysis stage will be possible.
* **Data Analysis for detecting systemic risk.** Given the fact that the objective is to detect systemic risk, an NCA would like to know the market participant that has the major number of different counterparties. Therefore, our proposal is that a network analysis through the use of graphs will be the most appropriate method to use for this objective.

The advantages of using a graph analysis are:

* + Can be easily implemented
  + The graphical output is user friendly
  + The analysis will not only show the market participants that concentrate the major part of counterparties, but also from this analysis can be inferred who are the liquidity providers, and the market sentiment. We consider that market sentiment is based on trust, if market players are positive and trust the markets, and the counterparties, then the number of trades and counterparties will be increased, otherwise if there is a market crisis, the trust between market participants will be reduced and the graph will immediately represent this situation, and as a consequence NCAs will be aware and will detect such a situation.

## 4.1 Data & extraction and transformation

We will assume that the NCAs will have the reports stored in a database, in order to illustrate our process, we are going to use a SQL Server database for extracting the data contained in the reports, and this will be the starting point of the data transformation pipeline.

In order to illustrate it, we will emulate the processes of data transformation and the implementation of the networks analysis as well.

The information of the trades will be extracted from different tables of the Trade Repository, we will run some queries and the results will be aggregated by pair of counterparties and by type of trade OTC or ETD.

The fields extracted are the following: Party\_ID, Counterparty\_ID, Venue of execution and the number of trades between the two counterparties.

The field Venue of execution will be the datum that will help us to distinguish between the ETD and OTC trades based on the following facts:

According to the technical standards of the regulation, the field Venue of execution should be informed as follows:

The field is an alphanumerical of four characters, and shall contain the values ‘XXXX’ or ‘XOFF’ for bilateral trades (over the counter) or in the case of reporting an Exchange Traded Derivative (ETD), the field shall be reported using a Market Identifier Code (MIC) included in the ISO 15022 specification.

In order to simplify our analysis, the trades reported with XXXX or XOFF will be marked as OTC, the trades with MIC codes will be classified as ETD.

Our proposal is to use an open source environment like Jupyter in order to run the processing through the use of some Python libraries.

We will create some scripts for extracting and transforming the data in order to finally use obtain a gml file that will be used for drawing and analyzing the graph.

We will use the following Python libraries:

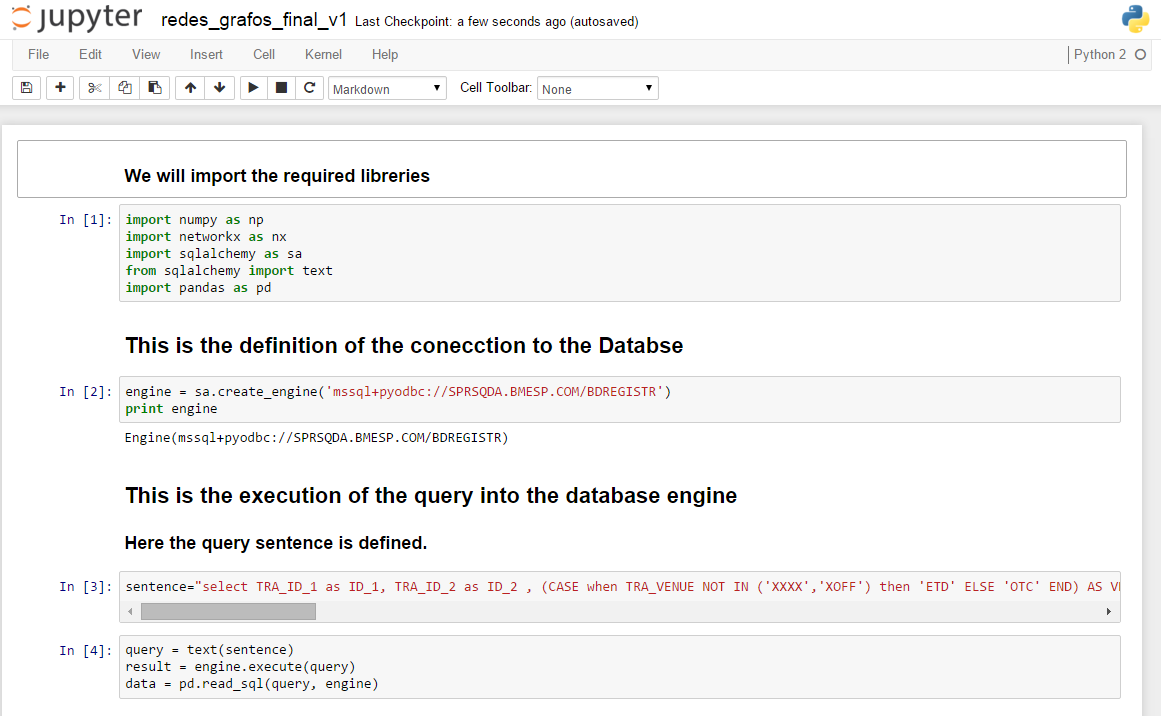
**Numpy**. This will be used for creating the graph.

**Networkx**. This library will enhance the process of graph generation and also this library produces a gml file.

**Sqlalchemy**. This library will be used for extracting the data from the Database and transform to a manageable format in Python.

**Pandas**. This library will be used for cleaning and transforming the data set.

Once we have produced the gml file, we will use Gephi for drawing the Graph and to obtain the analysis of the different values related to the Graph.

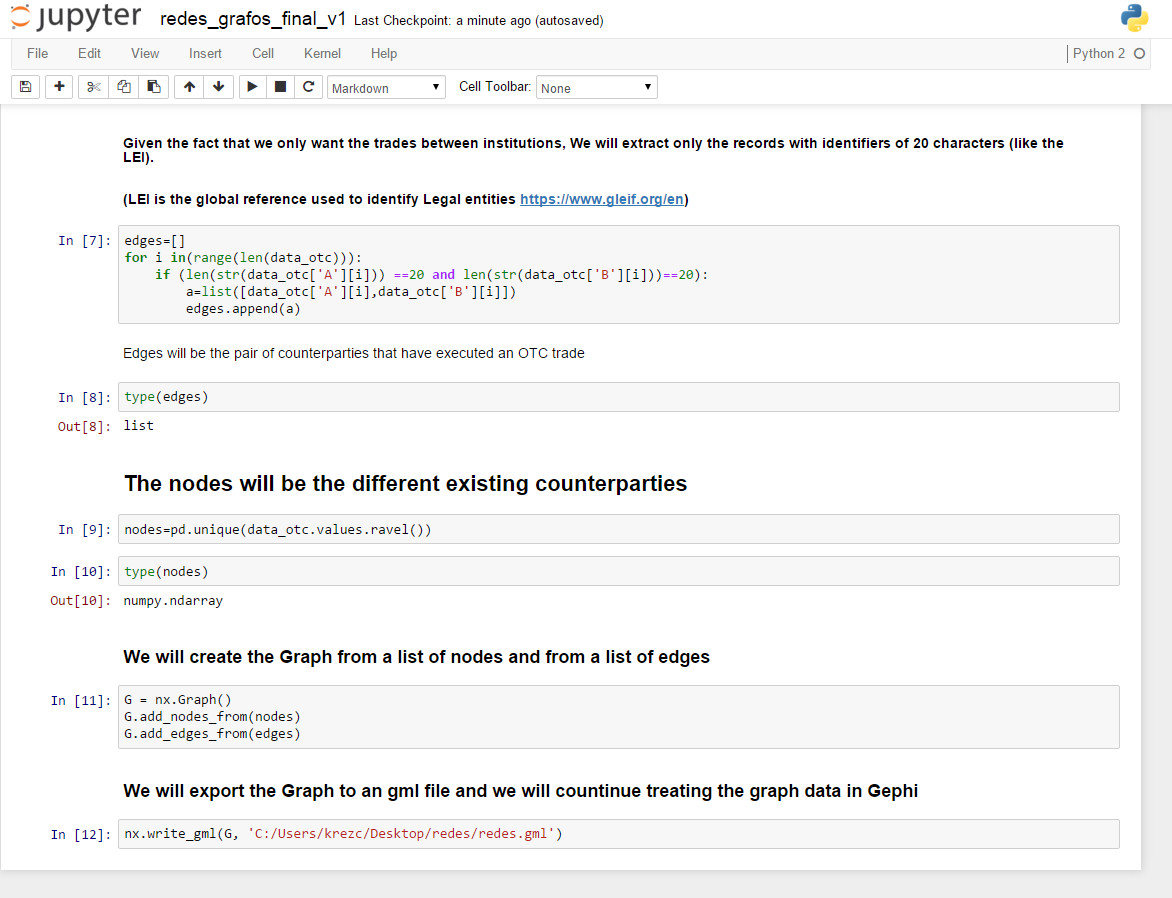
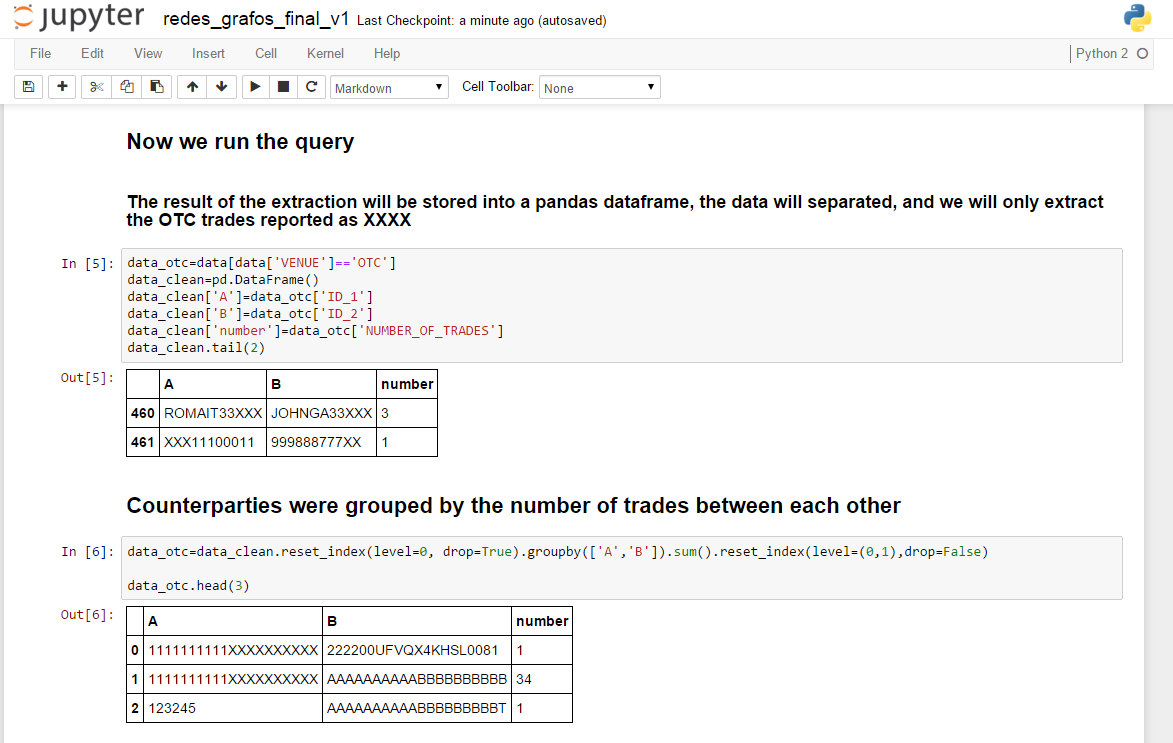


This is the definition of the SQL sentence that extracts the pairs of counterparties that have executed trades, we are selecting the two identifiers, and we are distinguishing the trades using an alias to the Venue of execution field, in addition to that we are excluding the trades where the other coutnerparty is an individual (natural person) or a non european counterparty, this is flagged by the field TRA\_ST\_REC INDI and NEEA accordingly.

The union all function was included so as to append the results of the two different tables extracted.

The maturity date or expiration date of the trade has been limited because we wanted to include the outstanding trades, because this is where the market risk is alive.

*select TRA\_ID\_1 as ID\_1, TRA\_ID\_2 as ID\_2 , (CASE when TRA\_VENUE NOT IN ('XXXX','XOFF') then 'ETD' ELSE 'OTC' END) AS VENUE, count(\*) as NUMBER\_OF\_TRADES from BDREPORTS where TRA\_ST\_REC NOT IN ('INDI','NEEA') AND TRA\_DATE\_MATU > '20151102’ group by TRA\_VENUE, TRA\_ID\_1, TRA\_ID\_2*



## Construction of the Network

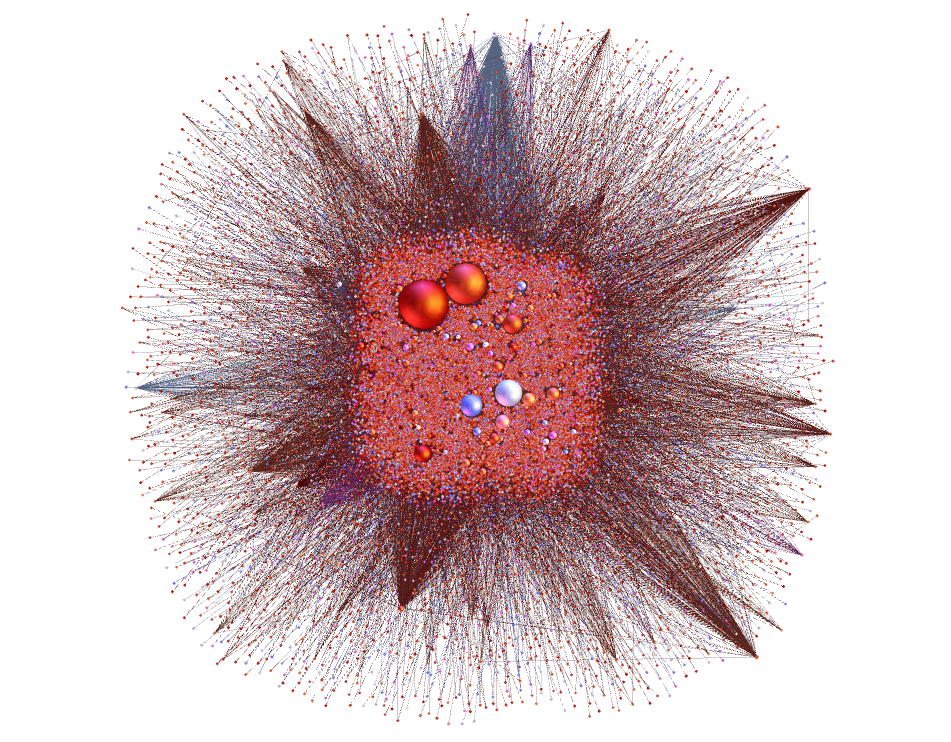
Once that the gml file has been produced with networkx library of Python, this library supports a wide range of imports and export formats, this GML file is a GraphML format that will be used in Gephi where the resultant graph will be analysed and a visualization will be produced.

In order to generate the graph, each counterparty has to be considered as a node in the network, and the edges will be the trading relationship between two counterparties, therefore the only relationship between two nodes is defined by the fact that a counterparty has whether or not executed a trade with other counterparty.

## 4.2 Data Analysis for detecting systemic risk

## 4.2.1 Network analysis

Once that the gml file has been loaded in Gephi, the Graph will be drawn by the program as shown below.



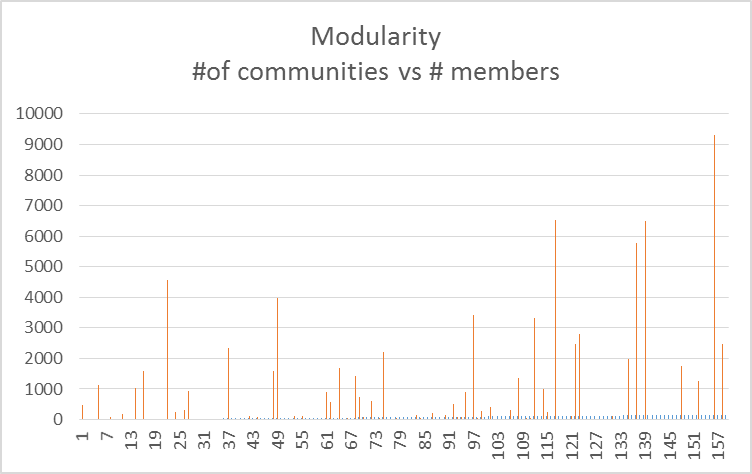
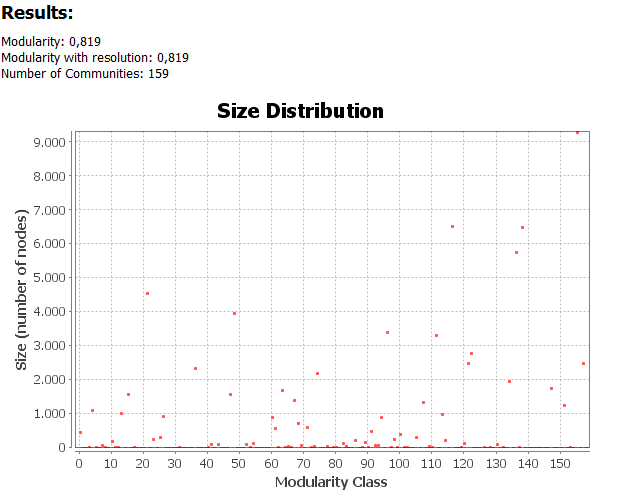
The network resulted in a graph of 81577 nodes (counterparties) and a total number of 111193 edges.

The Average Degree of the network resulted in 2.726

The distribution of the degree demonstrates that the major part of the connections is done by a couple of nodes.

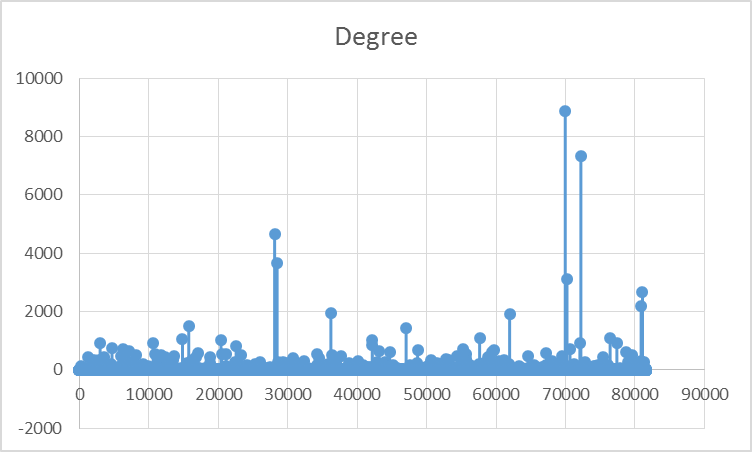
As mentioned before, this node analysis is providing a lot of useful information for the supervisory purposes, then an NCA will be able to understand the market sentiment, by comparing the different snapshots of the markets by visualizing at a macro level the existent relationships across market counterparties.

As part of the analysis, Gephi allows the user to obtain data certain metrics that give a descriptive characteristic of the drawn node like Modularity, Degree, Page Rank, and so on.



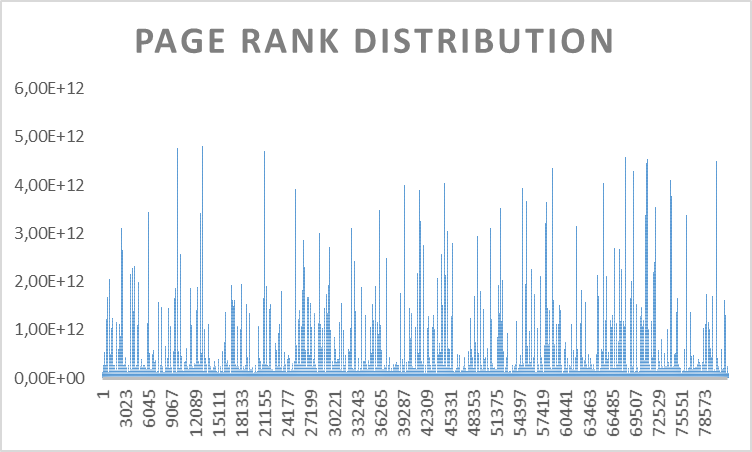
Modularity is 0.819

We obtained 159 Communities, our assumption is that the communities can be given by trading a specific asset class or a this can be explained by the types of counterparties that are trading i.e. by country, financial/no financial counterparties, etc.



Average degree: 2.726

Here we can se that just a few counterparties have a high Degree , therefore there are just a few counterpqarties that have the bulk of connections. Presumably this can be explained by the fact that these counterparties or systemic concentrators for instance this can represent the sell side of the trades.



The Page rank distribution appeared to be uniform, this can reflect a highly connected graph accross the entitites.

# 5 Conclusions

Our proposal intends to improve the efficiency of the supervisory tasks by European NCAs, during this document we have illustrated that the use of some technologies can facilitate and enhance different levels of analysis that might help NCAs to supervise and detect anomalies, market sentiment and risk concentrators.

Based on the fact that some NCAs do not have resources for implementing complex infrastructures, our proposal is based on open source tools that will enhance and facilitate their regulatory duties in a more efficient manner. The tools used for this methodology are Python as a programming language, Jupyter as an analysis interface, SQL server as a store device , and Gephi as a graph analysis tool, all of them are open source.

Other outcomes that can be taken from this methodology are conclusions like the confirmation of our initial hypothesis, this means that the bulk of the systemic risk is concentrated in just a small part of the counterparties, and the default or credit event of one of the entities will produce a “domino effect” to the OTC markets and as a consequence markets may face another market meltdown.

It is important to remark that the National Competent Authorities and the Financial Stability Board should have to focus their supervisory activities on these counterparties, otherwise the consequences on the OTC markets will be negative, and the supervisory efforts can be efficiently used.

An additional conclusion taken by the application of this methodology is based on the fact that trading is based on the existing trust across the different market players, therefore it would be useful for NCAs to store the different generated graphs drawn in a different periods, in order to understand how the market sentiment is evolving in the OTC markets.

Our methodology can be extended to other markets like the ETD, or the level of granularity could be increased - like breaking down the graph into different graphs by asset classes - or even this methodology can be complemented with other analyses like exposure analysis.

1. <http://ec.europa.eu/finance/financial-markets/derivatives/index_en.htm> [↑](#footnote-ref-1)