

# Social Network Analysis and Big Data tools applied to the Systemic Risk supervision

Mari-Carmen Mochón

*Ph program Uned University*

**Abstract** — After the financial crisis initiated in 2008, international market supervisors of the G20 agreed to reinforce their systemic risk supervisory duties. For this purpose, several regulatory reporting obligations were imposed to the market participants. As a consequence, millions of trade details are now available to National Competent Authorities on a daily basis. Traditional monitoring tools may not be capable of analyzing such volumes of data and extracting the relevant information, in order to identify the potential risks hidden behind the market. Big Data solutions currently applied to the Social Network Analysis (SNA), can be successfully applied the systemic risk supervision. This case of study proposes how relations established between the financial market participants could be analyzed, in order to identify risk of propagation and market behavior, without the necessity of expensive and demanding technical architectures.

**Keywords** — Social Network Analysis, Big Data, Financial Markets, OTC Derivatives, Systemic Risk, Supervision, Graphical analysis, EMIR, ESMA, G20, Trade Repository.

## I. INTRODUCTION

FOLLOWING the 2008 financial crisis, the G20 established an international forum for the heads of government of the world's major economies, to reach some consensus about the crash and put in place a high level plan to remediate those causes.

Since that moment, G20 jurisdictions started to draft and issue different regulatory reporting obligations, in order to provide greater transparency to the financial markets and supervisory capability to the national and international supervisors. In Europe several regulations, which imply a reporting obligation, have been implemented or are currently underway; EMIR [1], REMIT [2], SFTR [3] or MIFIR [4].

As a consequence, new market infrastructures appeared; trade repositories (TRs), to collect data from the industry, make sure of the reliability of the data, store it and make it available to regulators. Article 9 in EMIR (European Market Infrastructures Reform) mandates all counterparties to report details of any derivative contract they have concluded, or which they have modified or terminated, to a registered or recognized trade repository under the EMIR reporting requirements. TRs centrally collect and maintain the records of all derivative contracts.

OTC (Over the Counter) derivatives are financial products, which are negotiated bilaterally by two counterparties, and therefore are not registered in an official market. The OTC derivatives market, the financial engineering and the lack of transparency have played a key role in the crisis suffered in 2008.

Each of these regulations imply the daily reporting of billions of messages to be processed and stored in the trade repositories. And even though EMIR was the first regulation to impose a reporting obligation of the derivatives market to trade repositories in February 2014, further

reporting regulations appoint TRs as the data centers for the collection of the financial records of different segments of the industry.

Currently in Europe there are several trade repositories operating under EMIR regulation. They collect the information from the whole derivatives industry and make it available to European national competent authorities (NCAs) as defined in article 81 of EMIR.

All market agents are challenged by these regulatory obligations. Market participants have to develop and implement new reporting flows, trade repositories have to create complex and reliable technical architectures to ensure confidentiality and robustness, and at the end of the chain, regulators need to develop the relevant tools to be able to digest and understand the millions of pieces of information they have at their disposal.

If systemic risk supervisors do not achieve their objectives, all the previous work, effort and investments will be of no use. National and international supervisors are facing these mandates with limited human and technical resources, and therefore the achievement of this objective is not always easy.

## II. OBJECTIVE OF THE CASE OF STUDY

Regulators whose mandate is to monitor the systemic risk are now facing an adaptation process as hard as the one suffered by market participants. Even if supervisors already analyze an important amount of information, after the 2008 crisis and the G20 commitment towards transparency of the financial markets, millions of new records are under the scope of the regulatory reporting.

This case of study proposes complementary methods to monitor the systemic risk, making usage of the tools that Big Data provides. Big Data technical solutions can cope with the 3 main issues when analyzing and managing data, the 3 Vs; velocity, variety and volume. In the financial industry, the variety of the data is not a technical barrier, as reports are transmitted in quite standardized formats. On the other hand, volume and velocity suppose a big challenge for regulators.

Currently TRs receive hundreds of millions of transactions on a daily basis. National and international authorities are the data consumers, but still need to acquire the capability of extracting the information hidden behind the millions of data records contained in the trade repositories.

Analysis methodologies currently used for Social Network Analysis can be equally applied to the analysis of the relations established between market participants. This could provide valuable information related to the market behavior, tendencies and confidence in the market and clearly identify those participants who hold the highest propagation risk.

Social network analysis (SNA) is the analysis process of mapping and measuring the relationships, connections and exchanges taking place between people, agents, organizations, machines... The nodes in the network are the participants and groups while the links show relationships or communication exchanges between the nodes. SNA provides both a visual and a mathematical analysis of human, or business in this case, relationships.

### III. METHODOLOGY

#### A. Data subject to analysis & hypothesis

This case of study proposes the application of SNA methodologies and Big Data tools, currently used to analyze social networks to study the relations established between financial market participants and its tendencies within the OTC derivatives markets.

This exercise describes how graphical analysis tools could be applied to the OTC derivatives trades executed between European counterparties and reported to a trade repository under EMIR regulation. These reported trade details follow a predefined set of Reporting Technical Standards [5] (RTS) defined by ESMA (European Securities Markets Authority). Based on the reporting technical standards, it can be defined how such analysis should be designed.

Every derivative trade reported to a trade repository must include the reporting fields defined by ESMA in the reporting technical standards. Among the 85 fields subject to the reporting obligation, the following ones are the ones applicable to perform the proposed exercise; Reporting Counterparty ID (RTS field 1.3), ID of the Other Counterparty (RTS field 1.5), Venue of execution (RTS field 2.15).

The fields Reporting Counterparty ID and ID of the Other Counterparty represent a unique code identifying the counterparties to the transaction. The admitted values for this fields are; Legal Entity Identifier (LEI) (20 alphanumeric digits), Interim Entity Identifier (IEI) (20 alphanumeric digits), BIC (11 alphanumeric digits) or a client code (50 alphanumeric digits).

The field Venue of execution will be the datum used to distinguish between the ETD and OTC trades. According to the regulatory technical standards, the field Venue of execution should be informed as follows:

The field is alphanumeric of four characters, and shall contain the values 'XXXX' or 'XOFF' for bilateral trades (OTC) 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.

The trades reported with "XXXX" or "XOFF" will be marked as OTC, the trades with MIC codes will be classified as ETD.

In order to proceed with the analysis, the following assumptions have been considered:

- The UTI that identifies each trade is truly unique and commonly used by each pair of counterparties.
- All counterparties correctly populate the "venue of execution" field.
- Each regulator has at its disposal the information related to the derivative transactions subject to its supervisory mandate, sent by the TRs.

#### B. Data collection methodology

The data collection methodology used for this case of study is the "Extract Transform Load and Analyze" process (ETL&A).

According to article 81 of EMIR, all the NCAs will have access to the information reported to a TR. This proposal assumes that the information to be analysed by a Competent Authority will be contained in a flat file like a CSV.

In order to enhance the user interface, a framework like the Jupyter notebooks [6] is used in order to explore the data contained in the reports. These notebooks are a powerful open source tool that provide a user-friendly interface in the use of different Python libraries [7] like Pandas [8], Numpy [9] or Networkx [10]. Python is an open source programming language that allows quick and flexible integration of systems.

The proposed ETL&A process will be composed of the following phases:

- Data extraction: extract the data from a flat file to a Jupyter notebook.

- Transform: the data is selected and sliced according to the analysis requirements criteria.
  - A new data frame including the; Reporting Counterparty ID, ID of the Other Counterparty and Venue of execution is created.
  - The relevant transformations are applied in order to define the relations established between counterparties of the OTC derivatives market.
  - Networkx graph library is used in order to transform the data into a graph file.
- Load: load the graph file obtained into Gephi tool [11], an open source Social Network Analysis tool.
- Analyze: Gephi analysis and visualization tool provides outcomes and calculations that will illustrate the network and related metrics.

The below image represents an example of the script that should be created as part of the ETL process, using Jupiter Network\_Analysis.

```

We import the required libraries. I.e:
In [1] import pandas as pd

*The user can define the value of the variable input_path with the path where the file is stored

*The result of the data transformation will produce an output file. The user needs to define where the file will be stored
In [2] input_path= "PATH/WHERE/THE/FILE/IS Report_in_a_flat_format.csv"
output_path= "PATH/WHERE/THE/FILE/WILL/BE/GENERATED"

A plain file like a csv can be read from the python notebook
*The pandas library will create a DataFrame
*The csv will be stored in a Pandas DataFrame by the use of the read_csv function
*This function will include the input_path, the columns to be selected and the separator. I.e:
In [3] data=pd.read_csv(input_path,usecols=['Counterparty ID, ID of the other Counterparty, Venue of Execution'])

*The data contained in a DataFrame can be selected based on a condition
*In this case we can extract the records with n value in field Venue of Execution
In [4] We select the records where the field venue is XXXX. I.e:
data_otc=data[data['VENUE'] == 'XXXX']

Out[4]:


|     |                      |                      |
|-----|----------------------|----------------------|
| 111 | LEIEX178911234567890 | LEIEX378911234567890 |
| 112 | LEIEX278911234567890 | LEIEX478911234567890 |



The construction of the graph requires the creation of the nodes and edges
In [5] Edges=[]
for i in range(len(data_otc)):
    if (len(str(data_otc['A'][i]))==20 and len(str(data_otc['B'][i]))==20):
        a=list(data_otc['A'][i],data_otc['B'][i])
        edges.append(a)
The nodes will be the existing counterparties
In [6] nodes=pd.unique(data_otc.values.ravel())

Finally the Graph can be created with the nodes and edges that were previously defined
In [7] G=nx.Graph()
G.add_nodes_from(nodes)
G.add_edges_from(edges)

In [8] A GML File can be exported to perform a Network analysis with a tool like Gephi.
To export the gml file to the output path:
nx.write_gml(G,output_path)

```

Fig.1. ETL process script - Source: Own elaboration

#### C. Tools and programs applied to analyze the data

Gephi is a visualization and exploration software for all kinds of graphs and networks. It is an open-source free tool commonly used in the social networks analysis such as Twitter or Facebook, as provides simple information of complicated networks created by the billions of participants interacting via internet. The software uses a 3D render engine to display large networks in real-time and to speed up the exploration.

Once a graph file is generated and loaded in the Gephi application, the program draws a network that reflects the relations of all the different parties. Additional metrics are generated automatically by the application.

Social Network Analysis can be described as the process of investigating social connections, interactions and structures, by using network and graph theories [12]. The tools used for applying this

technique enhance the visualization of sociograms in which nodes are represented as points or entities and the interactions can be represented by bridges or lines. Such visualization has drawn significant attention in the recent years, because this technique helps researchers to understand, find and predict patterns, and interactions among social actors, i.e., identifying central actors, roles, subgroups or clusters.

The Gephi statistics and metrics framework offer the most common metrics for social network analysis (SNA) and scale-free networks:

- Betweenness Centrality, Closeness, Diameter, Clustering Coefficient, PageRank
- Community detection (Modularity)
- Random generators
- Shortest path

#### D. Social Network Analysis metrics

The following are some of the most relevant metrics used in SNA:

- Degree Centrality

Social network analysts measure the network activity of a determined node by using the concept of degrees, this is the number of direct connections a node has. The nodes holding the most direct connections in the network, are the most active nodes and therefore the 'connectors' or 'hubs' in the network.

- Betweenness Centrality

The location a node has in the network also determines the importance that node has in the network. A node might not have many direct connections, but if it has a good location, close to other important nodes, it may have a powerful 'broker' role in the network. On the other hand this nodes may imply a single point of failure. Supervisors must keep an eye on those nodes with high betweenness, as they has great influence over what flows, and does not, in the network. Location is a key element in network analysis.

- Closeness Centrality

Another important metric is the "closeness centrality" meaning how quickly a node contact any other one. I node may not have many direct and indirect ties, but still have access to all the nodes in the network more quickly than anyone else. These nodes have the shortest paths to all others, meaning that they are close to everyone else. These nodes are in an excellent position to monitor the information flow in the network, as they have the best visibility into what is happening in the network.

- Network Centralization

The relationship between the centralities of all nodes can reveal much about the overall network structure. A very centralized network is dominated by one or a few very central nodes. If these nodes are removed or damaged, the network quickly fragments into unconnected sub-networks. A highly central node can become a single point of failure or financial crash in that market. A network centralized around other set of well-connected hubs can find the turnaround if that hub is disabled or removed. Hubs are nodes with high degree and betweenness centrality.

The less centralized a network is established, the less single points of failure it has. Healthy and robust markets should not be very centralized in order to provide alternatives in the market in case of an agent making default in their financial liabilities.

#### IV. RESULTS

The graph below is an example of the graphical illustration a regulator could obtain following the proposed methodology to analyze the connections established by the different market participants of a given target sample. Each circle would represent a market participant. Whenever a party is connected to a higher number of other market

participants, the circle is represented bigger. This is measured through the "degree". The degree of a node is the number of relation (edge) it has.



Fig.2. Example Network graph result – Source: Gephi 0.9.0

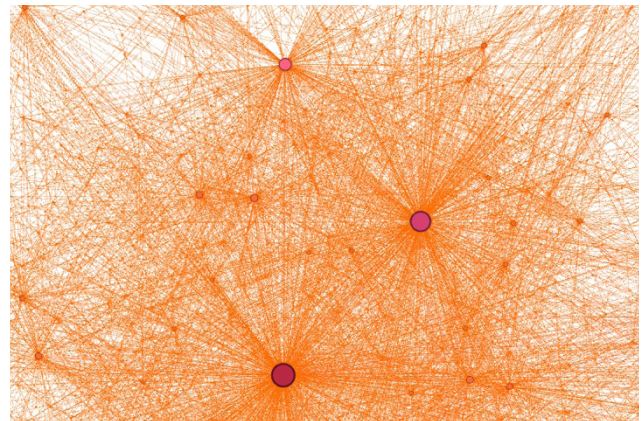


Fig.3. Example Network graph result – Source: Gephi 0.9.0

This graph shows a highly connected network where only a few nodes are significantly bigger than the others. These nodes have the highest number of connections established in the network. Consequently, they would represent a higher propagation risk. Should a credit event occur to any of these nodes, it can be assumed that an important size of the market would be affected.

The layout of the graph distributes the nodes within the graph depending on the closest relations and communities identified.

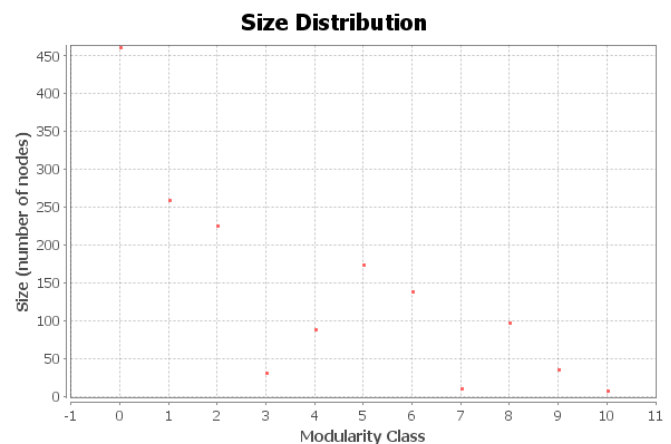


Fig.4. Example of System Modularity representation – Source: Gephi 0.9.0



In addition to the main graph that represents all the relations established in the network, the Gephi application generates the most common metrics for social network analysis and scale-free networks.

For instance, the modularity of the given example is reflected in Fig.5. This system is made up of relatively independent but interlocking communities of members.

The degree of interconnection between counterparties is also identified. A few nodes have a high degree. This would mean that just a few counterparties have the bulk of connections. Presumably, those counterparties or systemic risk concentrators, represent the sell side of the trades, providing financial services to many market participants.

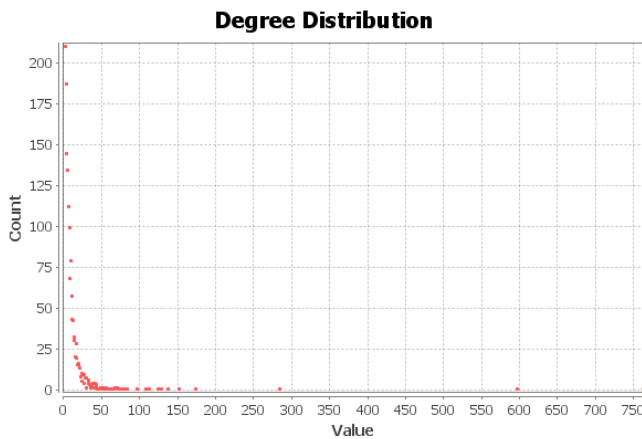


Fig.5. Example System Degree – Source: Gephi 0.9.0

## V. CONCLUSIONS

Financial markets are clear examples of social structures whose behavior is dictated by external factors. Macroeconomic or microeconomics events, expectations, speculation or risk appetite are some of the factors that compose a market sentiment which is expressed through the commercial relationships established among themselves. As stated in the article “The Digital Economy: Social Interaction Technologies” [13]; “*the daily activities of many businesses are being socialized*”. Consequently, these connections and interactions can be analyzed just as any other social network.

The proposed approach may be used to analyze market behavior, tendencies and confidence feelings between counterparties by studying the relations established market participants. Additionally, the identification of the relations established between market participants can disclose information regarding propagation risk factors, and potential cascading failure situations. Through a simple process of analysis, it is possible to identify who are the market participants highly trusted and active in their respective financial fields. These participants concentrate the highest propagation risk and in consequence are systemic risk concentrators.

When many participants establish relations with different counterparties, confidence reigns in the market, and participants are not very sensible to the counterparty credit risk. On the other hand, when the number of relations decrease and only few participants are trusted in the market, it can be interpreted that the market is suspicious and already anticipating potential credit defaults.

Additionally, by analyzing the evolution of the network interaction, it is possible to monitor market feelings and confidence in the market. Competent authorities can monitor market sentiment, by comparing the different snapshots of the market network, by visualizing at a macro level the existent relationships across market counterparties. The

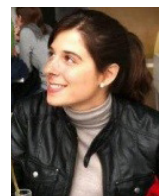
analysis tool offers dynamic graph analysis, where users can visualize how a network evolve over time by manipulating the embedded timeline.

It is also important to point out that the tools used in this analysis are free and open-source, therefore no initial cost, license fee or important technical architecture investment is required. Still, these powerful analytic tools enable supervisors to extract relevant information. Regulators with budget constraints can contemplate these analytical tools as an option.

This is a complementary analysis that could be used by regulators in addition to the traditional analysis already performed. Further and more complex studies could be performed with these and other Big Data tools.

## REFERENCES

- [1] European Markets Infrastructure Regulation <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32012R0648>
- [2] Regulation Energy Market Integrity and Transparency <https://www.acerem-it.eu/portal/public-documentation>
- [3] Securities Financing Transaction Regulation <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R2365>
- [4] Markets in Financial Instruments Regulation <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014R0600>
- [5] ESMA Reporting Technical Standards <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:052:0001:0010:EN:PDF>
- [6] Jupyter notebook <http://jupyter.org/>
- [7] Python standard library <https://docs.python.org/2/library/>
- [8] Pandas documentation: <http://pandas.pydata.org/pandas-docs/version/0.17.1/>
- [9] Numpy documentation <http://docs.scipy.org/doc/>
- [10] Networkx documentation <https://networkx.github.io/documentation/latest/>
- [11] Gephi documentation <https://gephi.org/users/>
- [12] Otte, Evelien; Rousseau, Ronald (2002). “Social network analysis: a powerful strategy, also for the information sciences”
- [13] Teófilo Redondo “The Digital Economy: Social Interaction Technologies”. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 3, N° 2. 2015



**Mari-Carmen Mochón**, Institutional Relationship Manager, REGIS-TR. She was born in Sevilla in 1983. In 2006 she was licensed in Business Administration by the University of CUNEF in Madrid and a master title in “Financial Markets and Alternative Investment” in the Instituto BME in 2008. She is Specialized in Regulatory reporting systems/compliance, Institutional and Government Relations, OTC Derivatives products & Financial Markets. She has worked for the Spanish Derivatives Exchange supervision area until 2009. After she worked at the Spanish Central Bank for the European Central Bank, in the creation of a European securities settlement platform. In 2011 she moved to REGIS-TR, for the creation project of a European Trade Repository owned by the Deutsche Börse Group & Bolsas y Mercados Españoles. Currently she is the Institutional Relationship Manager, being responsible for International and European government affairs as well as Market Institutions. She has combined her professional career in the financial sector with an academic path; University teacher at Universidad Pontificia de Salamanca (UPSAM), teaching “International Economy” and “Economy of the Europe Union”. She has also being post-grade program teacher, in the course “Economic crimes and prevention of financial risk”, at the National University for Distance Education (UNED). She has several publications such as; “Teoría de las organizaciones”, F. Mochón, M.C. Mochón y M. Sáez, 2015 Alfaomega, Mexico, “Gestión organizacional”, F. Mochón, M.C. Mochón y M. Sáez, 2015 Alfaomega, Mexico, “Administración”, F. Mochón, M.C. Mochón, M. Sáez, 2014. Alfaomega, Mexico and “Economía” in a team of three, 2002. McGraw Hill, Spain. Currently she is finalizing a master degree in Business Intelligence and Big Data in Colegio Universitario Cardenal Cisneros, Madrid, and it is enrolled in the doctoral program at the Faculty of Economics, UNED University.