UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE POLITICHE, ECONOMICHE E SOCIALI

M.Sc. in Data Science and Economics

SOCIAL NETWORK ANALYSIS ON POST TRADE DATA: CENTRALITY, SCALE FREE BEHAVIOR AND RESILIENCY OF THE NETWORKS OVER TIME

Supervisor:
Prof. Giancarlo MANZI
External Co-Supervisor:
Dr. Enrico PAPALINI

Candidate:
Andrea Ierardi
960188

Academic Year 2020/2021

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

The aim of financial integration in Europe, is to create a single European financial market and to extend growth of Members States by generating new investment opportunities and strengthening the EU banking sector. For this purpose, European Central Banks (ECB) launched in 2008 the TARGET2-Securities (T2S) platform for cross-border trading that started operations in 2015. T2S is a technical answer to support Central Securities Depositories (CSD) by providing borderless and neutral settlement services [1].
Monte Titoli is an Italian Central Securities Depository (CSD) and an organization of the Euronext group. It is the only Italian CSD authorized to operate securities settlement services which it does via the Express II system[1]. Express II combines a clearing or netting system and a gross settlement system. Monte Titoli operates under internationally common rules and standards and has an automated procedure for determining the optimal settlement result given the participants' holdings of securities and cash. A Central securities depository (CSD) is a specialized monetary organization holding securities like shares, allowing an easy transfer of ownership through a book entry rather than by a transfer of physical certificates. This allows brokers and financial companies to hold their securities at one location where they can be available for clearing and settlement. Settlement operations are executed through the T2S platform, that allows the settlement of instructions between two participants

(intra CSD) or one CSD's participant and another CSD's participant (cross CSD settlement). T2S each day has several operations and trades that follows a strict schedule [2]. The system has long daily operating hours, opening at 7 *a.m.* CET and closing at 6 *p.m.* CET.
During the day time, the day trade phase starts from 7 *a.m.* and 6 *p.m.*. To allow participants to better manage their end-of-day liquidity, customer payments are subject to a cut-off time of 5 *p.m.*, while for bank to bank payments the cut-off is at 6 *p.m.*. It starts the new business day on the evening of the previous day at 7 *p.m.*, a "night-time window" being available from 7.30 *p.m.* to 7 *a.m.*, with a three-hour technical maintenance period between 10 *p.m.* and 1 *a.m.*. The night-time window facilitates the night-time settlement of various systems in central bank.
Another important institution is Cassa di Compensazione e Garanzia and it plays a central role of central counterparty. A central counterparty[3] is an institution interposed in securities trades between the two contracting parties, protecting the latter against default risk and guaranteeing a successful execution. The central counterparty protects itself against its own risk by taking securities or cash collateral proportionate with the value and risk. Only one central counterparty is authorized in Italy, Cassa di compensazione e garanzia (CC&G)[3]. Originally, CC&G dealt only with financial derivatives, but with time its activities were extended to shares and Italian government securities, while now it covers a wide range of trading platforms and financial instruments.
During a collaboration with Monte Titoli we have the opportunity to work with post-trade data. In particular, this data includes many important and famous clients such as banks, large-sized companies but also small companies. The research question of this thesis deals with the study of post-trade processes between main companies of Monte Titoli. In particular, the post-trade settlement instructions may follow some characteristics and behaviors that we want to understand and analyse. We are interest in knowing how several companies trade between each other, the cash flow of these trades and which companies play a central role in the system. For this purpose, Social Networks Analysis provides

intuitions into macroscopic and microscopic behaviors of different actors in the networks, providing useful insights in the topological structure.
Social network Analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. SNA[4] theoretical roots come from the work of early sociologists in the late 1800s, Émile Durkheim and Ferdinand Tonnies. Tönnies argued that social groups can exist as personal and direct social ties that either link individuals who share values and instrumental social links. Durkheim stated that social phenomena arise when interacting individuals constitute a reality that can no longer be accounted for in terms of the properties of individual actors. Georg Simmel essays pointed to the nature of network size on interaction and to the likelihood of interaction in ramified, distant networks rather than groups[5]. Social scientists have used the concept of social networks since early in the 20^{th} century to define complex sets of relationships between members of social systems[6]. SNA has found applications in various academic disciplines such as biology[7], communication, economics, geography, history, political science[8], public health[9] and computer science[10], as well as practical applications such as countering money laundering[11] and terrorism. Furthermore, SNA is formally referred to as a set of actors, or nodes that are connected by one or more types of relations, called links or edges. Nodes are the units that are connected by a relation links. The units are most commonly individuals or groups that can be connected to other units. Examples of nodes are actors, individuals, web pages, blogs, emails, messages, companies, families, cities, positions, or nations.
The most fundamental global properties of graphs are degree-based measures. The various edges connecting two points form a path whose length can be defined by the number of edges that chain together to form it. Movement through a network depends importantly on the path distances involved. Thus, the communication of ideas or innovations may occur through a whole chain of intermediaries with little or no reliance on the participants all having direct personal knowledge of each other. This was the key to Milgram's discussion of the 'small world'[12]. Measures of path

distance must consider the directionality and value of the various edges. A whole collection of global measures gather the idea of the centrality of points within their graphs. The degree of a knot is the most basic measure and has been defined as local centrality. Locally central points are well-connected within particular parts of the network but may not be at all well-connected in a global sense. The most commonly used measure of closeness involves calculating the aggregate distances from each point to all others.

SNA properties and utilities can be exploited for answering our research questions. Thus, graphs provide a way of representing relationships mathematically. Each trade can be abstracted as a relationship established by two or more companies. Post post-trade data can be represented by graph in which companies are nodes and edges are the trades. Cash flow traded can be represented by an edge where source and destination nodes are characterized by two companies. This solution leads in the simplification of the large amount of daily data in Monte Titoli systems. Degree distribution of such graph describes the most important and central nodes. Furthermore, other topological measurements like average clustering, assortativity, transitivity and path length allow an in-dept analysis of the networks structure. Exploiting the degree distribution leads in the understanding of a behavior and a patter in the structure of the networks. These are the so called scale-free network in which the degree distribution follows a power law. Understanding whether a network is scale-free helps define a model that can describe the underlying structure of the network, including growth behaviors. For our purpose, this can help study the behavior of companies in the system over time.

Monte Titoli has been merged with other Euronext CSDs and renamed Euronext Securities. Furthermore, Cassa Compensazione e Garanzia has been renamed in Euronext Clearing[13]. The previous Italian names will be retained in the following description referring to Euronext Securities and Euronext Clearing.

Chapter 1 introduces the concept of Central securities depository (CSD), Monte Titoli and the key role of the T2S platform. Furthermore, we introduce the concept of Social Network Analysis (SNA) and the possibility of graphs construction from post-trade data.
Chapter 2 shows the literature review and describes several studies. The first study is about scale-free networks degree behavioral in social networks analysis. There are different degree patterns of scale-free networks: super-weak, weakest, weak, strong, strongest, and not scale-free. Another study explains the topological analysis of the economic structure of an entire country based on Swedban payments data. Moreover, a focus on change in topology after central and random node elimination. Another work is about networks' topology formed by the payment flows over a Fedwire Funds Service. A in-dept analysis on topology changes after disruptive events is performed.
Chapter 3 presents the experiments and results of the analysis on Monte Titoli post-trade data. We define how data is extracted from T2S system and stored in Amazon Web Service (AWS). In particular, the preprocessing pipeline of the data for social networks construction. Social networks graph are generated cumulative monthly, non-cumulative monthly and daily. For monthly and daily data different experiments are presented. On monthly networks an initial data inspection is performed on main topological characteristics: mean degree, assortativity, average clustering, transitivity density, connected components. Central nodes are identified and at what frequency during the period from May 2018 through July 2021. Moreover, different degree of scale-free behavioral are detected in the networks and different results are obtained for cumulative and non-cumulative networks. Finally, an in-depth study of the resilience of networks after the elimination of central and random nodes is performed. Two case studies of disruptive events are considered for daily networks: the impact of Covid19 and BTP emissions on networks topology.
Chapter 4 presents final remarks and future works.

Chapter 2

Literature Review

In this section we will present several studies regarding social network analysis that will be referred to in subsequent sections. The article "Scale-free networks are rare"[14] describes how social networks may present scale-free patterns and to what degree. The paper *"On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia"*[15] explains the main topological characteristic in the scale-free networks from Swedbank and it describes an analysis on central and random nodes elimination. In the article *"The Topology of Interbank Payment Flows"*[16] shows how networks topology changes after disruptive events such as the terrorist attacks of September 11^{th}, 2001. The study *"Comparison of Failures and Attacks on Random and Scale-Free Networks"*[17] shows how resiliency is higher in scale-free than random networks in case of failures, while scale-free are more vulnerable to localized attacks.

2.1 Scale-free networks are rare

The article *"Scale-free networks are rare"*[14] shows how scale-free networks rarely appear in real-world sets of data. Several types of Scale-free degree networks are described:

- **Super-Weak**: for at least 50% of graphs, no alternative distribution is favored over the power law.

The second type represents direct evidence of scale-free structure, and the various modifications of a scale-free pattern can be organized in a set of nested categories that represent increasing levels of evidence:

- **Weakest**: for at least 50% of graphs, a power-law distribution cannot be rejected.
- **Weak**: requirements of Weakest, and the power-law region contains at least 50 nodes.
- **Strong**: requirements of Super-Weak, Weakest and Weak and $2 < \hat{\alpha} < 3$ for at least 50% of graphs. Where $\hat{\alpha}$ is the scaling parameter.
- **Strongest**: requirements of Strong for at least 90% of graphs, and requirements of Super-Weak for at least 95% of graphs.
- **Not Scale Free**: Networks that are neither Super-Weak nor Weakest.

In the study, they evaluated the degree distributions of nearly a thousand real-world networks from a wide range of scientific domains and found that scale-free networks are not ubiquitous. Fewer than 36 networks exhibited the strongest level of evidence for scale-free structure, in which every degree distribution associated with a network is convincingly scale free. Only 29% of networks exhibited the weakest form and for 46% of networks, the power-law form was not necessarily itself a good model of the degree distribution, but was simply a statistically better model than alternatives. 49% of networks showed no evidence of scale-free structure, and in 88% of networks, a log-normal fits the degree distribution as well as or better than a power law. The results demonstrate that scale-free networks are not a ubiquitous phenomenon.

2.2 On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia

The paper *"On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia"*[15] presents a topological analysis of the economic structure of an entire country based on payments data obtained from Swedbank. Data is based on around 80% of Estonia's bank transactions and it is represented as social networks where nodes are customers of the bank and the links are established by payments. In addition, they study scaling-free patterns presence and structural properties of this network such us its topology and components. They show that this network shares typical structural characteristics known in other complex networks: degree distributions follow a power law, low clustering coefficient and low average shortest path length. They identify the key nodes of the network and perform simulations of resiliency against random and targeted attacks with two different approaches: random and hub nodes removal. In the study emerges that scale-free networks are more resistant to random disconnection of nodes since one can eliminate a considerable number of nodes randomly and the network's structure is preserved and will not break into disconnected clusters. However, the error tolerance is acquired at the expense of survival attack capability. When the hubs (most connected nodes) are targeted, the diameter of a scale-free network increases and the network breaks into isolated clusters. One of the characteristics that makes a hub or a key node an important one, is its high betweenness not just its high degree. Hubs connect sub-areas of the graph that are not connected to one another directly. In fact, most influential nodes are critical because they shorten path lengths by making reachability high and information moving fast. In the study more than of thousand of key nodes were found, this means that around 8% of the nodes are relevant to keep the structure connected. If these nodes are removed, then the number of connected components and average path lengths between the nodes would increase, leaving the network vulnerable to breakage.

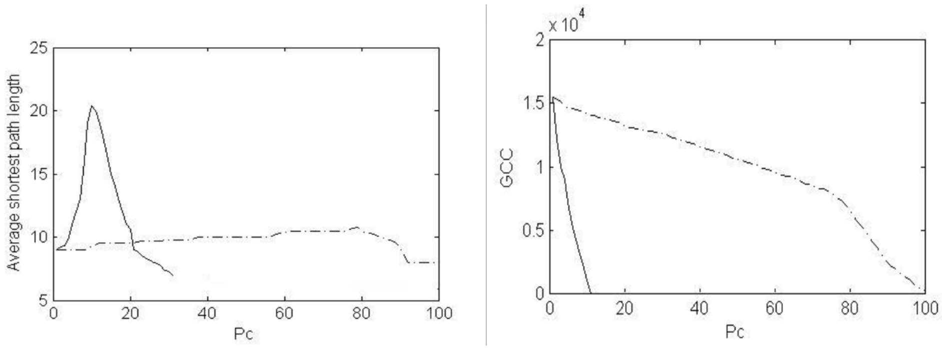


Figure 2.1: Image of the targeted and random damage over the network of payments. (a) The average shortest-path length in the GCC against the percentage of deleted nodes. (b) The GCC against the percentage of deleted nodes. Continuous lines display the effect of the targeted removal and the dashed lines display the effect of the random removal of nodes. Pc is the percolation thresholds.

2.3 The Topology of Interbank Payment Flows

In the article *"The Topology of Interbank Payment Flows"*[16], they analyzed the topology of about 60 daily networks formed by the payment flows between commercial banks over a Funds Service, called Fedwire. These networks share many of the characteristics commonly found in other empirical complex networks, such as a scale-free degree distribution, high clustering coefficient and the small world phenomenon. The networks are disassortative like many other technological networks. They find that networks are compact despite low connectivity and they include a tightly connected core of money-center banks to which all other banks connect. The degree distribution is scale-free over a substantial range. Moreover, they find that the properties of the network changed considerably in the immediate after-effects of the attacks of September 11, 2001. The topology of the networks were significantly altered by the attacks of September 11^{th}, 2001 as shown in Figure 2.2. The number of nodes and links in the network and its connectivity was reduced, while the average path length between nodes was significantly increased. The performance of scale-free networks under ordinary and disrupted conditions is receiving increasing attention. Scale-free networks, for example, have been found to preserve their connectivity under random node removal yet to be vulnerable to disconnection following removal of high-degree nodes.

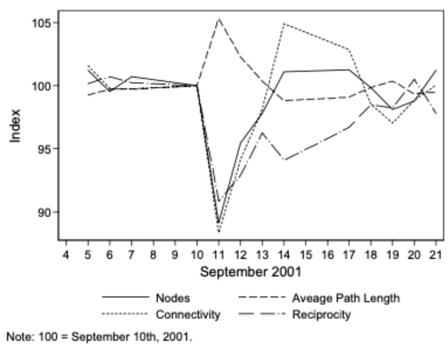


Figure 2.2: Size, average path length, connectivity and reciprocity Sept. 4 - 21, 2001, Sept. 10, 2001 = 100

2.4 Comparison of Failures and Attacks on Random and Scale-Free Networks

The study *"Comparison of Failures and Attacks on Random and Scale-Free Networks"*[17] shows how scale-free networks resulted more robust than random networks in case of failures, while they are more sensitive to localized attacks. The resilience of the network to failures or attacks can then be analyzed by studying how the size of the connected component varies as a function of the number of removed nodes (or links). For random networks, the size of the largest connected component drops to zero when a finite fraction of the nodes are removed, while for scale-free networks, it decreases slowly, and reaches 0 only when most nodes have been removed. They demonstrate that scale-free networks appear to be much more resilient to failures than random networks. Moreover, they propose several types of attack strategy on social networks. An attack strategy is the way the nodes (or links) are chosen during an attack. The classical attack consists in removing nodes by decreasing order of their degree. They also provide two new attack types: almost-failures attack that consists in randomly removing nodes of degree at least 2 and efficient link attack where links are randomly removed between nodes of degree at least 2.

Chapter 3

Experiments and results

The aim of the project is the construction of Social Networks based on post-trade data. Two ways of constructing networks are defined: monthly and daily networks. On monthly networks are performed topological analysis, central nodes detection, analysis on scale-free patterns and resiliency after random and hub deletion. On daily networks two case studies of disruptive events are examined: the impact of Covid-19 on network metrics and topology and the impact of large emission of BTP Italia and BTP Futura.

3.1 Data Extraction

T2S (TARGET2-Securities) is a European securities settlement engine which aims to offer centralized delivery-versus-payment (DvP) settlement in central bank across all European securities markets[18]. Since the large number of daily settlement instructions in T2S system, data are retrieved using Python Spark library[19] and stored in a specific Simple Storage Service (S3) bucket in Amazon Web Services (AWS). AWS[20] is a on-demand cloud computing platforms and APIs that provides a variety of basic abstract technical infrastructure and distributed computing building blocks and tools. Moreover, SageMaker service and computational power are used for the developing of the following analysis. The settlement instructions extracted for analysis range from May 2018 through

July 2021.

Company name	
Adventure Solutions Electronics	South Construction
Contract Bell	Net Venture
Federated Speed Adventure	Application Data Resource
Direct Analysis	West Net Innovation
Analysis Venture Industries	Research Systems Design
People Graphics Star	Advanced Resource Hill
Vision Provider	Graphics General
Speed Graphics Resource	Design Studio
Digital East	Network Interactive Telecom
Star Bell	Internet People Internet
Omega Vision Graphics	East Analysis
North Signal Atlantic	Consulting Digital
Hardware Provider Pacific	Contract Future Internet
Solutions Consulting West	Hardware Electronics Power
Advanced Resource Venture	Data Telecom Star
Advanced Design Analysis	Contract East
Adventure North	Building Direct Galaxy
Direct Studio Vision	Innovation Adventure Data
Solutions Construction Solutions	Atlantic Research Application
Data North	West Omega
Architecture East Construction	Electronics Venture Federated
Interactive Source Medicine	Provider Source Solutions
Resource Galaxy	Universal Vision
Bell Virtual Internet	Signal Solutions

Table 3.1: Example of company names

For privacy reasons, Monte Titoli's customers are anonymized using a random name generator with the exception of Monte Titoli and Cassa Compensazione e Garanzia. Table 3.1 shows an example of randomly generated company names. From a set parts of common firm names such as "Solutions", "Construction", "Electronics" etc., we randomly combine them two or more times and we obtain a random string name. As the strings are generated, the association between the random names and the companies is stored in an internal and private file to maintain a fixed terminology and to be consulted later. Since the random generator takes

a list of words as input, randomly combines them, and produces a name, none of the outputted names refer to a particular company name or feature. Furthermore, codes that might hint at which company is being referred to have been blacked out.
The dataset extracted from T2S comprehends different features: International Securities Identification Number (ISIN), Electronic Documentation Center (EDC), Italian banking association (ABI), descriptive name of the company, source company name, destination company name, source Bank Identifier Code (BIC), destination Bank Identifier Code (BIC), Settlement Account (SAC) T2S, cash flow value, settlement status, type of financial instrument and Exchange Traded Funds (ETF) indicator, Cross Border Instruction Indicator.

3.2 Data Pre-processing and graphs definition

Pre-processing techniques are applied on the dataset to better extract and manipulate the information. In particular, data are aggregated by month and by business day. Specifically, since each business day a company can trade multiple times with the same company, there could be conflicts during edge generation phase. To solve this problem, the cash flow traded by each company is aggregated by the couple receiver and sender for both daily and monthly data. In this way there are not duplicated edges. Furthermore, instruction settlement status is filtered by settled (S) and failed (N) while partial (P) is not considered. The reason why partial instructions are excluded is because an instruction from partial state on a specific business day, would end up in settled or failed state on subsequent days. Another reason is that partial instructions are not common. In addition, free of payment instructions are excluded because the cash flow is not valued. A large amount of data can be abstracted and managed in a better way by using graphs. In fact, social network structures are useful for process simplification, visualization and inspection of large amount of data. The strategy consists in constructing graphs in which nodes represent companies and links represent exchanges of instructions. Links are weighted using the instruction cash flows. For network construction and topological analysis, python library called NetworkX [21] is used. There are several instructions during a business day and the aggregation are done using the following criteria:

- Daily aggregation: for each business day, cash flow values are aggregated by sender and receiver name and instruction status.
- Monthly aggregation: for each business day, cash flow values are aggregated by sender and receiver name, instruction status, financial instrument type. Each month is independent of the previous one (Figure 3.2).
- Monthly cumulative aggregation: same aggregation as monthly networks, however in this case nodes and links are added each month. For cash flows, instead, the valued are cumulative summed each months. Each month depends on the previous one (Figure 3.1).

For each combination of type aggregation and feature, a different network is constructed. In fact, each network has different nodes, links, and characteristics. Algorithm 1 shows the pseudo-code for monthly networks construction, while Algorithm 2 shows the pseudo-code for daily networks.

Algorithm 1 Monthly Network construction algorithm

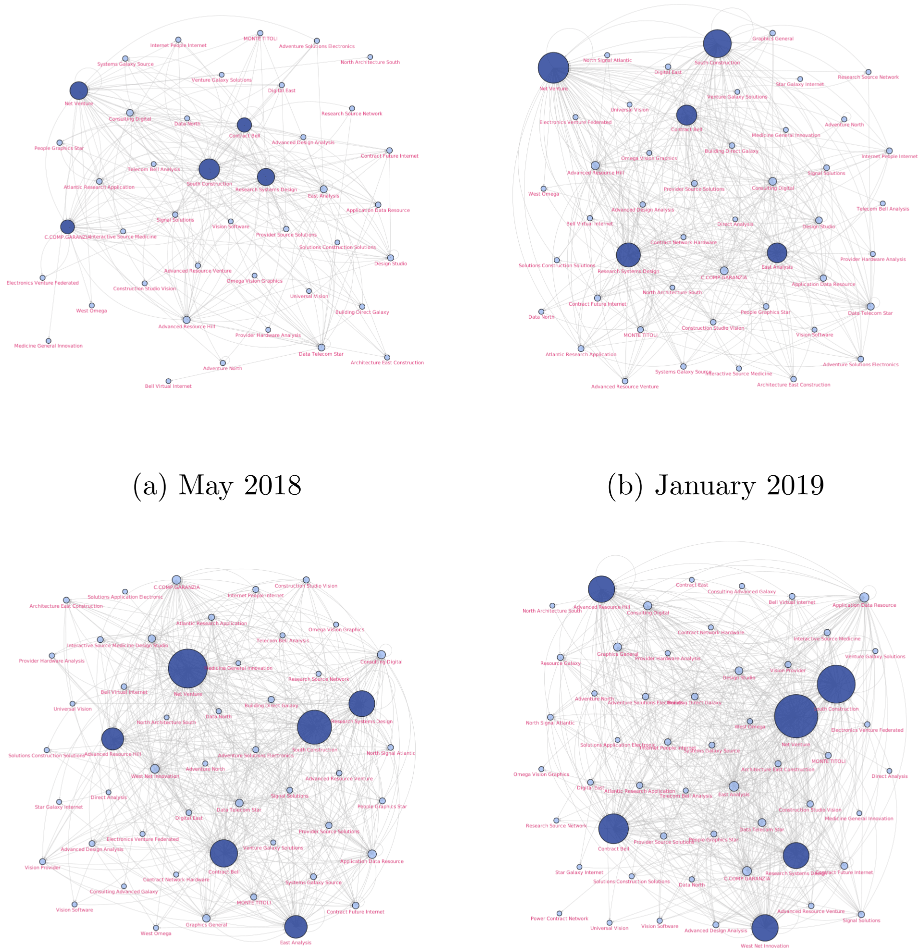
Require: $data\_dict$ set of sets containing months, instrument types and instruction status
  $G\_dict \leftarrow \emptyset$
  for $month$ in $data\_dict$ **do**
    $G\_dict[month] \leftarrow \emptyset$
    for $instr$ in $data\_dict[month]$ **do**
      $G\_dict[month][instr] \leftarrow \emptyset$
      for $status$ in $data\_dict[month][instr]$ **do**
        $G\_dict[month][instr] \leftarrow DirectedGraph()$
        $source\_list \leftarrow data\_dict[month][instr][status][source]$
        $dest\_list \leftarrow data\_dict[month][instr][status][dest]$
        $w\_list \leftarrow data\_dict[month][instr][status][w]$
        for $source$, $dest$, w in zip($source\_list$,$dest\_list$, w_list) **do**
          add $edge$ $(source, dest, w)$ **to** $G\_dict[month][instr][status]$
        end for
      end for
    end for
  end for

Algorithm 2 Daily Network construction algorithm

Require: $data\_dict$ set of sets containing months and instruction status
  $G\_dict \leftarrow \emptyset$
  for day in $data\_dict$ **do**
    $G\_dict[day] \leftarrow \emptyset$
    for $status$ in $data\_dict[day]$ **do**
      $G\_dict[day][status] \leftarrow DirectedGraph()$
      $source\_list \leftarrow data\_dict[day][status][source]$
      $dest\_list \leftarrow data\_dict[day][status][dest]$
      $w\_list \leftarrow data\_dict[day][status][w]$
      for $source$, $dest$, w in zip($source\_list$,$dest\_list$, w_list) **do**
        add $edge$ $(source, dest, w)$ **to** $G\_dict[day][status]$
      end for
    end for
  end for

The financial instrument types considered are:

- Government Bonds: a government bond is a debt security issued by a government to support government spending and obligations[22].
- Corporate Bonds and similar: a corporate bond is a type of debt security that is issued by a firm and sold to investors[23].
- Shares and similar: shares are units of equity ownership in a corporation[24].
- Mutual funds: mutual funds consist of a portfolio of stocks, bonds, or other securities. They are operated by professional money managers, who allocate the assets and try to produce capital gains or income for the investors[25].
- Exchange-Traded Fund (ETF): is a basket of securities that trades on an exchange just like a stocks[26]. ETF share is bought and sold all day, while mutual funds only trade once a day after the market closes. ETF can contain all types of investments, including stocks, commodities, or bonds. Figure 3.1 and Figure 3.2.
- Others: financial instruments not included in the above categories.



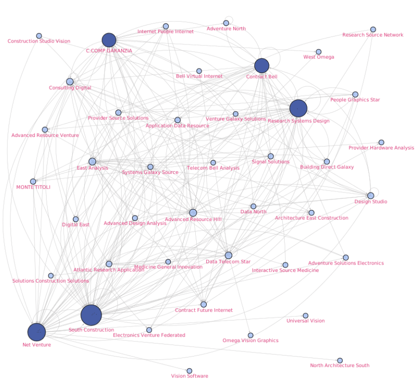
(a) May 2018

(b) January 2019

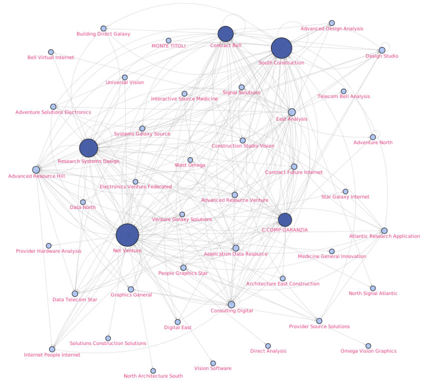
(c) January 2020

(d) July 2021

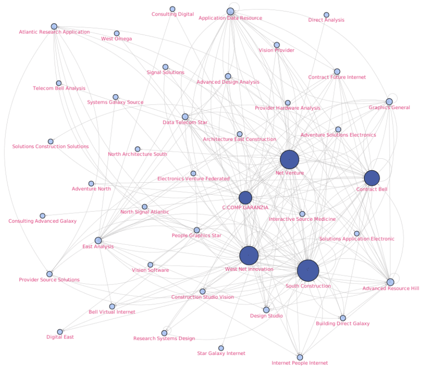
Figure 3.1: Cumulative ETF network with only-failed instructions growing over time.



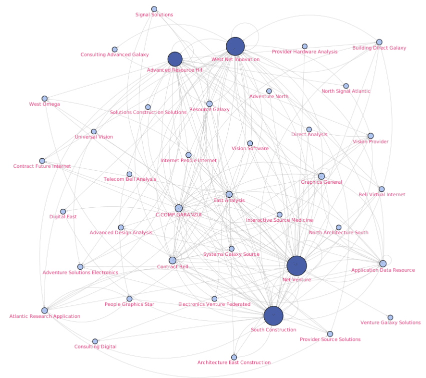
(a) May 2018



(b) January 2019



(c) January 2020



(d) July 2021

Figure 3.2: Non-cumulative ETF network with only-failed instructions growing over time.

3.3 Monthly Graph

Several analysis are performed on monthly graphs. An inspectional analysis on the overall structure of the networks, with a focus on topological measure such as number of nodes, links, mean degree, assortativity, average clustering, transitivity and density, weakly and strongly connected components. A centrality analysis, detecting important nodes and the frequency over time. Scale-free networks pattern detection and their scale-free degree to understand how nodes and links are distributed in a network. A Networks resiliency analysis, focusing on the topology changes after the elimination of random and localized central nodes.

3.3.1 Networks topology

Nodes are the number of firms in a network, while links are the number of trades between companies. Size/order ratio, represents the number of trades per company:

$$ratio = \frac{links}{nodes}$$

	Nodes	Links	Size/Order ratio
ETF_S	54	711	13.167
ETF_N	54	534	9.889
Gov. Bonds_S	72	1311	18.208
Gov. Bonds_N	61	701	11.492
Shares_S	114	1194	10.474
Shares_N	60	726	12.1
Corp. Bonds_S	346	1312	3.792
Corp. Bonds_N	55	462	8.4
Funds_S	67	754	11.254
Funds_N	54	543	10.056
Other_S	56	448	8
Other_N	30	125	4.167

Table 3.2: Summary table representing overall number of links and nodes for networks for each different financial instrument types

Table 3.2 shows summary statistics of monthly aggregated graphs sepa-

rated by the different types of financial instrument. ETF networks with only-failed and only-settled instructions, have an equal number of companies but different number of links. Only-settled corporate bonds and similar, have more than triple the number of links and nodes than those only-failed. In addition, corporate bonds graph has the greatest number of nodes and connections. Shares only-settle network, on the other hand, has nearly doubled the number of connections compared to the network only-failed network. The reason why only-settled graphs are more connected and populated is that, in general, 90% of instructions result in a settled state, while failed instructions are present but rarer.
Other interesting metrics are assortativity, average clustering, transitivity and density. Assortativity describes how nodes in the network connect to each other based on their number of connections.[27]. This measure quantifies the tendency of nodes to be connected to similar nodes in a complex network, where similarity in this case is the degree of the node (i.e. the number of connections). This means a network is assortative if nodes with a certain degree are more likely to connect with nodes of similar degree, while a network is disassortative when nodes with low degree are more likely to be linked to those with high degree, and vice versa[28]. The clustering coefficient of a node is the tendency to cluster. It is the density around a specific node and represents the proportion of the closest nodes that are connected to each other [15]. The overall clustering coefficient is the average of the clustering coefficients. It indicates if there is a link between two nodes that have a common neighbor. Transitivity is the overall probability for the network to have interconnected adjacent nodes, thus revealing the existence of closely connected communities (or clusters, subgroups, cliques). It is calculated by the ratio between the number of observed closed triplets and the maximum possible number of closed triplets in the graph. Complex networks and notably small-world networks often have a high transitivity[29]. Density is the ratio of the number of existing connections to the maximum number of connections. A high network density implies many connections, and nodes in such networks have a higher chance of being interconnected[30].

	Assortativity	Average Clustering	Transitivity	Density
ETF_S	-0.484	0.778	0.465	0.248
ETF_N	-0.504	0.739	0.368	0.187
Gov. $Bonds_S$	-0.514	0.703	0.539	0.256
Gov. $Bonds_N$	-0.55	0.739	0.386	0.192
$Shares_S$	-0.506	0.406	0.537	0.093
$Shares_N$	-0.474	0.735	0.457	0.205
Corp. $Bonds_S$	-0.677	0.128	0.124	0.011
Corp. $Bonds_N$	-0.41	0.683	0.374	0.156
$Funds_S$	-0.483	0.619	0.456	0.171
$Funds_N$	-0.497	0.753	0.368	0.19
$Other_S$	-0.413	0.649	0.518	0.145
$Other_N$	-0.232	0.434	0.375	0.144

Table 3.3: Summary table representing the overall network assortativity, average clustering and transitivity for each financial instrument types

Table 3.3 shows a summary of the aggregate graph measures for each type of financial instrument: assortativity, average clustering, transitivity and density. The majority of the financial instruments exhibit an assortativity coefficient negative and smaller than -0.4, apart for $Other_N$. This means that in the networks, firms with numerous connections tend to connect more to those with a low number of links. Corporate $Bonds_S$ shows the greatest disassortativity with a coefficient of -0.677. Networks mostly show clustering coefficient values above 0.4. This means that companies tend to cluster very frequently. Furthermore, Corporate $Bonds_S$ networks show a lower average clustering coefficient with a value of 0.128. Networks exhibit a moderately high transitivity coefficient with values above 0.4. Corporate $Bonds_S$, as with clustering, show the lowest value for transitivity. The density is moderate in the networks and does not reach a coefficient greater than 0.248. This means that firms in the networks have sparse connections. ETF_S networks have the highest density with a score of 0.248, while Corporate $Bonds_S$ have the lowest density of 0.011, which is close to 0.
Networks connections are essential to determine connected components.

Given a directed graph, a weakly connected component (WCC) is a subgraph of the original graph in which all vertices are connected by some path, ignoring the direction of the edges. A strongly connected component (SCC) is the portion of a directed graph in which there is a path from each vertex to another vertex. If the graph is not connected, it can be broken down into several connected components.

	Weakly CC	Weakly connected?	Strongly CC	Strongly connected?	Node Isolated
ETF_S	54	Yes	54	Yes	0
ETF_N	54	Yes	49	No	0
Gov. Bonds_S	72	Yes	64	No	0
Gov. Bonds_N	61	Yes	51	No	0
Shares_S	114	Yes	65	No	0
Shares_N	60	Yes	55	No	0
Corp. Bonds_S	346	Yes	68	No	0
Corp. Bonds_N	55	Yes	44	No	0
Funds_S	67	Yes	58	No	0
Funds_N	54	Yes	49	No	0
Other_S	56	Yes	47	No	0
Other_N	30	Yes	18	No	0

Table 3.4: Summary table representing weakly or strongly connected, the number of nodes in the main component and isolated nodes

Table 3.4 shows weakly and strongly connected component patterns for graph aggregated by financial instrument types. The totality of networks are weakly connected, which means there that are no isolated nodes. Moreover, the number of weakly connected components changes between only-settled and only-failed networks. The only financial instrument type with equal number of weakly connected components in only-settled and only-failed is the ETF. The number of strongly and weakly connected components, on the other hand, differs between only-settle and only-failed networks for the other financial instrument types. The only strongly connected graph is the ETF_S. Interestingly, Corporate Bonds_S have 346 weakly connected components, while only 68 are strongly connected.

3.3.2 Central nodes

The average degree of a network is the number of links divided by the number of nodes. Considering a system of banks and financial institutions, the networks mostly have few connections per node, while only few "hubs" have thousands of links [16]. Two important node characteristics in a directed network are out-degree and in-degree. In-degree is the number of links that originate in the node, while out-degree is the number of links that terminate in the node.

	Mean degree	Mean In degree	Mean Out degree
ETF_S	26.333	13.167	13.167
ETF_N	19.778	9.889	9.889
Gov. $Bonds_S$	36.417	18.208	18.208
Gov. $Bonds_N$	22.984	11.492	11.492
$Shares_S$	20.947	10.474	10.474
$Shares_N$	24.200	12.1	12.1
Corp. $Bonds_S$	7.584	3.792	3.792
Corp. $Bonds_N$	16.8	8.4	8.4
$Funds_S$	22.507	11.254	11.254
$Funds_N$	20.111	10.056	10.056
$Other_S$	16	8	8
$Other_N$	8.333	4.167	4.167

Table 3.5: Summary table representing the overall network mean degree, in-degree and out-degree for each financial instrument types

Figure 3.5 shows the mean degree for each financial instrument type network for only-settled and only-failed instructions. Government bonds network with only-settled instructions has the highest degree in the system, while corporate bonds network with only-settled instructions has fewer connections per node. The mean in-degree and out-degree result equal, since on average for every incoming connection from one company there will be an outgoing connection in another.
Degree centrality assigns an importance score based simply on the number of links held by each node. In this case, it can be useful to look at both in and out degrees since we are considering transactional data[31].

Centrality is important because it indicates which node takes up critical position in a network. Central positions get equated with remarkable leadership, good popularity or excellent reputation in a network.
There are different measures of centrality:

- Degree Centrality: measure that counts how many neighbors a node has[32].
- Betweenness Centrality: measure how a one node is undertaking mediation role in a network. It measures the number of times a node functions as an intermediary bridge connecting two other nodes that do not have a direct connection[32].
- Closeness centrality: measure one node to the others nodes' sum distances. It reflects the distance from a node to the rest.
- Eigenvector centrality: measures node's importance while giving consideration to the importance of its neighbors[33].
- Page rank: PageRank considers the number of in-bound links, the PageRank of the linkers, and the link propensity of the linkers[34].

From a single network, a centrality score can be obtained for each node, in order to obtain a rank score for each company. However, obtaining a rank for each month from May 2018 to July 2021 could not bring to a consistent and unique solution. To overcome this problem, the frequency of each company's appearance in the top ten most central is calculated for each network. Thus, the summation of the number of appearance in the rank represents how many times nodes are central over time. Frequency can be a satisfactory indicator that summarizes the centrality of a company during the period from May 2018 to July 2021. The centrality algorithm considered for the rank score, is the PageRank weighted by cash flow.

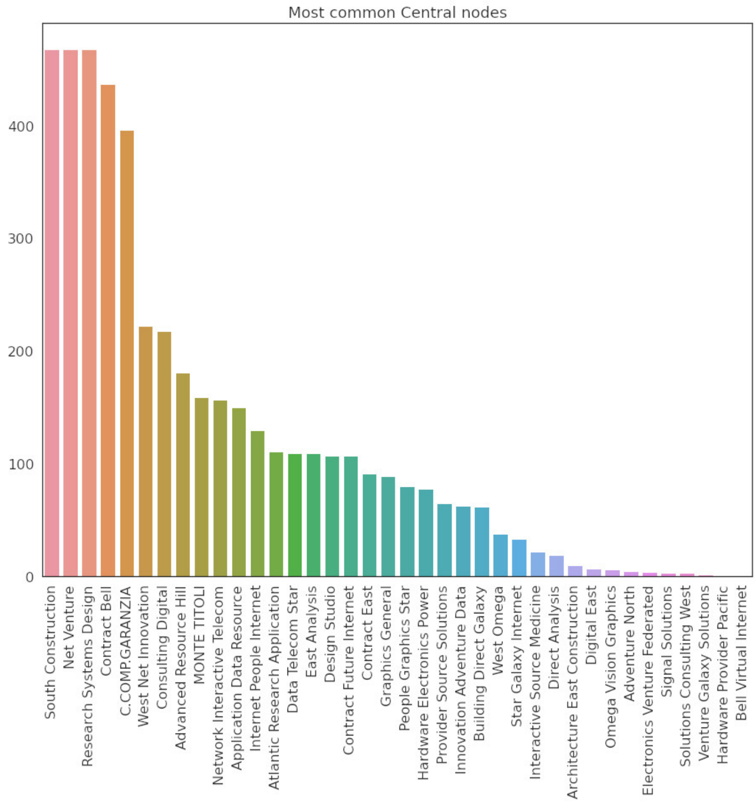


Figure 3.3: Most frequent central nodes for cumulative monthly networks

Considering monthly cumulative networks, Figure 3.3 shows that *South Construction*, *Net Venture*, *Research System Design* are the most frequently appearing nodes in the top-10 most central companies with a number of occurrences greater than 400. In the following positions of the ranking we have *Contract Bell*, *Cassa Compensazione e Garanzia* (*CC&G*) and *West Net Innovation*. CC&G has a high centrality because is the Italian entity deputed to clearing and netting.

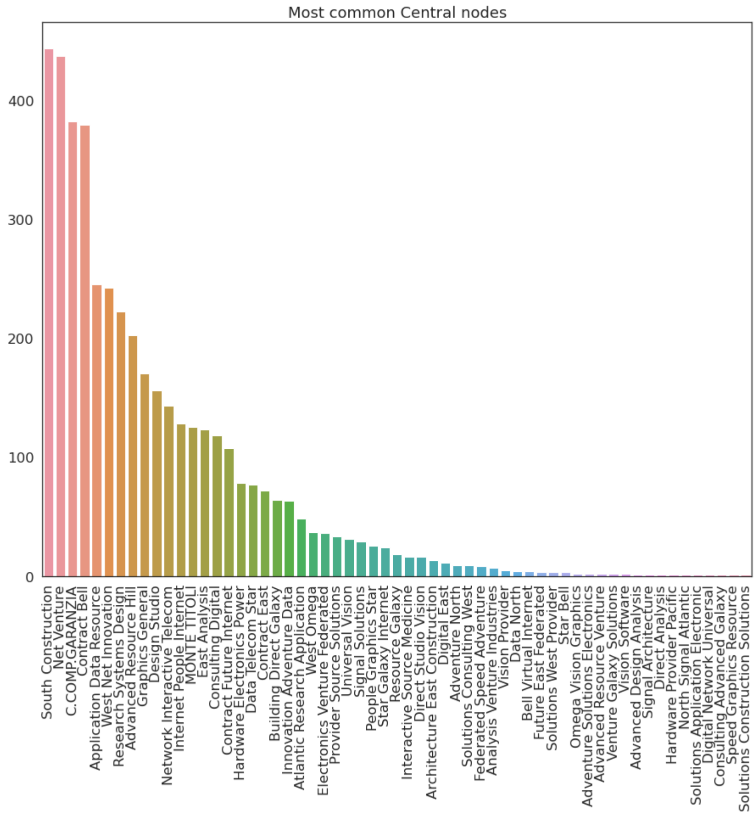


Figure 3.4: Most frequent central nodes for non-cumulative monthly networks

Considering monthly non-cumulative networks, Figure 3.4 shows that *South Construction*, *Net Venture* and *Cassa Compensazione e Garanzia* (*CC&G*) are the most frequently appearing nodes in the top-10 most central companies with a number of occurrences close to 400. In the following positions of the ranking we have *Contract Bell*, *Application Data Resource*, *West Net Innovation* with a number of occurrences close to 200.

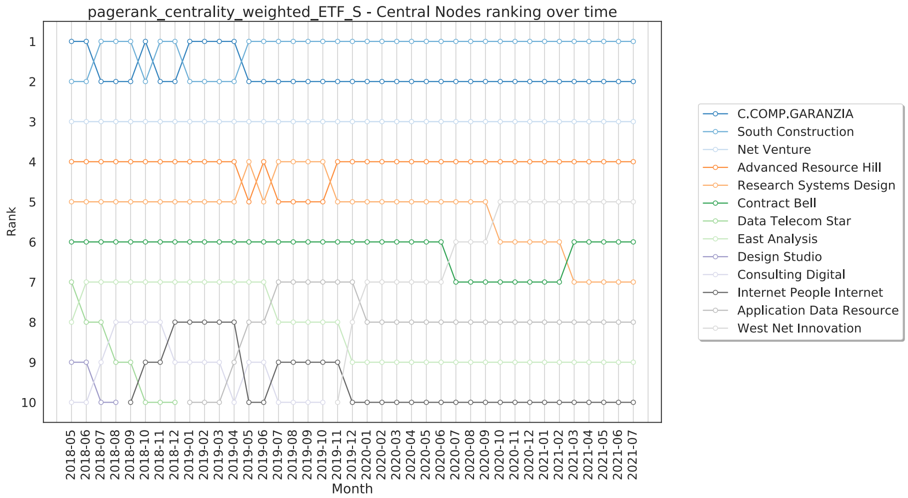


Figure 3.5: Page Rank Top-10 most central nodes over time for cumulative monthly ETF networks with only-settled instructions.

An in-depth monthly analysis is conducted to understand the change in the rank for several companies over time. Figure 3.5 shows the top-10 most central nodes over time for cumulative monthly ETF networks with only-settled instructions according to Page Rank. Initially, *CC&G* and *South Construction* companies continually exchange between the first and second ranking positions, while as the networks grow cumulatively, the positions stabilized. In the lower positions of the rank, instead, the companies hardly maintain the same position in the during years 2018 and 2019. The positions remain constant after 2019.

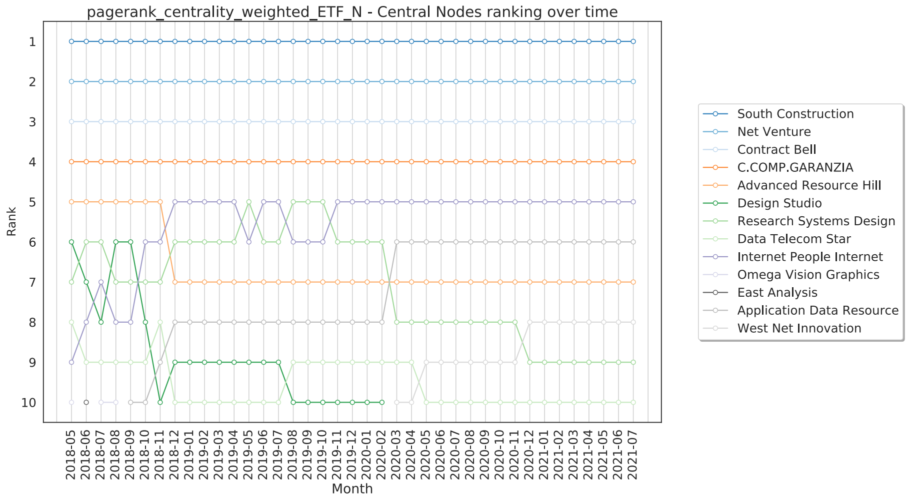


Figure 3.6: Page Rank Top-10 most central nodes over time for cumulative monthly ETF networks with only-failed instructions.

Figure 3.6 shows the top-10 most central nodes over time for cumulative monthly ETF networks with only-failed instructions according to Page Rank. *South Construction*, *Net Venture*, *Contract Bell* and *CC&G* maintain their positions throughout the time frame. In the lower positions of the rank, instead, the companies hardly maintain the same position.

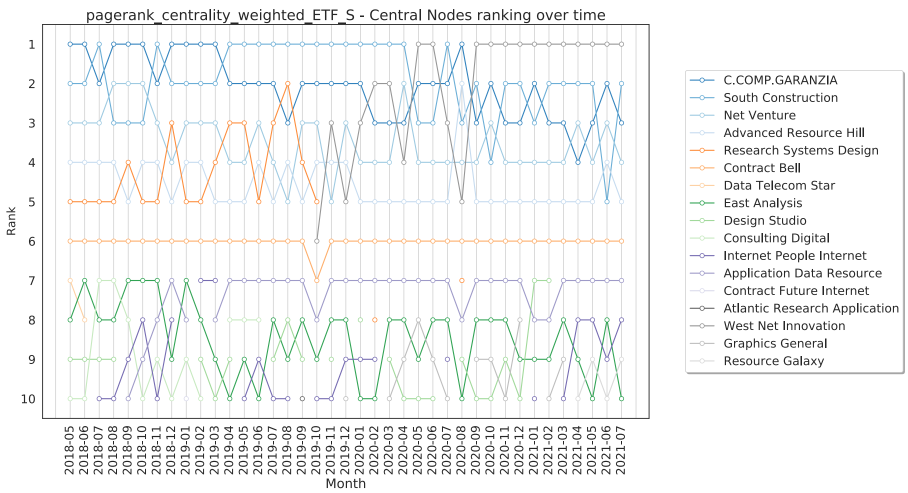


Figure 3.7: Page Rank Top-10 most central nodes over time for non-cumulative monthly ETF networks with only-settled instructions.

Figure 3.7 shows the top-10 most central nodes over time for non-cumulative monthly ETF networks with only-settled instructions according to Page

Rank. *CC&G* and *South Construction* started at first and second place, however, as time went on, their positions continued to fluctuate. *Contract Bell* is the most stable company, changing position only once in October 2019.

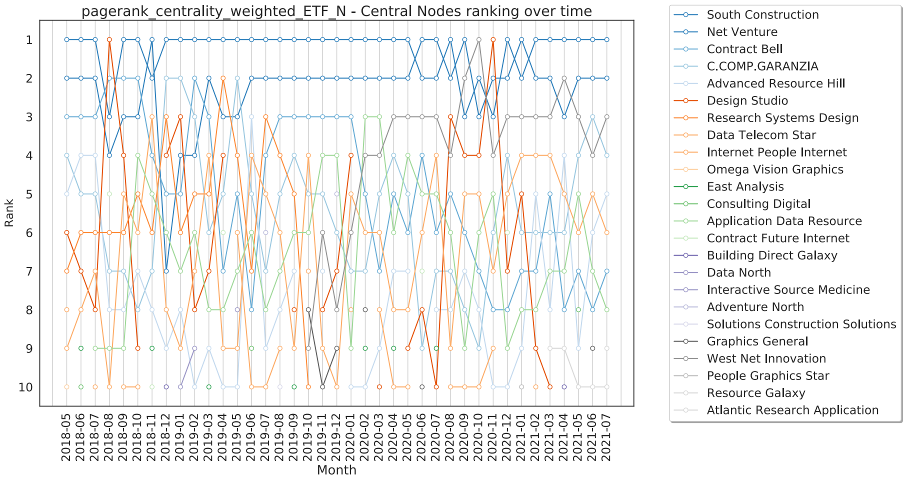


Figure 3.8: Page Rank Top-10 most central nodes over time for non-cumulative monthly ETF networks with only-failed instructions.

Figure 3.8 shows the top-10 most central nodes over time for non-cumulative monthly ETF networks with only-failed instructions according to Page Rank. There is not a stable pattern in companies ranking.

South Construction, Net Venture, Research System Design are the most frequently appearing nodes in the top-10 most central companie for cumulative networks, while South Construction, Net Venture and Cassa Compensazione e Garanzia (CC&G) are the most frequently appearing nodes in the case of non-cumulative networks. Any entity participating in a regulated market can partner with CC&G, which assumes the risk of counterparty default by becoming the counterparty in the contract itself. For this reason, expected that CC&G would exhibit a high centrality and it is confirmed from both cumulative and non-cumulative networks. Analysing the centrality over time, for ETF non-cumulative networks we can see a non-stable pattern of firms' positions over time. For ETF cumulative, instead, companies ranking is quite stable. In addition, only-failed networks exhibit a more unsteady behavior with respect to those only-settled.

3.3.3 Scale-free Networks

A network is *scale-free* if the characteristics of the network are independent of the size of the network (i.e. the number of nodes). This means that when the network grows, the underlying structure remains the same. A scale-free network is defined by the distribution of the number of edges of the nodes following a so called *power law* distribution. An example of central concern for macroeconomics are production networks whose scale-free nature has recently been put forward[35] as a potentially major driver of macroeconomic fluctuations. The fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as

$$P(K) \sim k^{-\alpha}$$

where α is a parameter whose value is typically in the range $2 < \alpha < 3$. Moreover, scale-free networks refer to small world theory which is based on the idea that two individuals will be connected through a short series of intermediaries. In the 1960s, Stanley Milgram tested this theory [12] in which all nodes are distant from each other for a short path. The "Six degrees of separation" is the idea that people on average are six, or fewer, social connections away from each other. As a result, a chain of "friend of a friend" statements can be made to connect any two people in a maximum of six steps.

Scale-free networks exhibit three main properties:

- Growth: a growth process where, over an extended period of time, new nodes join an already existing system.
- Preferential Attachment: in a network new nodes prefer to link to the more connected nodes.
- Scale-free Resiliency: scale-free networks are more resistant to random disconnection of nodes [17]. A considerable number of nodes

can be eliminated randomly while preserving the network's structure without breaking into disconnected clusters. When the most connected nodes are targeted, the diameter of a scale-free network increases and the network breaks into isolated clusters. This occurs because when removing these nodes, the damage disrupts the core of the system, whereas a random attack most likely does not.

A power-law distribution is also sometimes called a scale-free distribution and networks with such a degree distribution are referred to as scale-free networks. In section 2.1 strongly scale-free networks are considered rare, while in real world most scale-free networks exhibit a weak scale-free degree. We use sfanalysis [36] code to test if the networks considered are scale-free and to what degree.

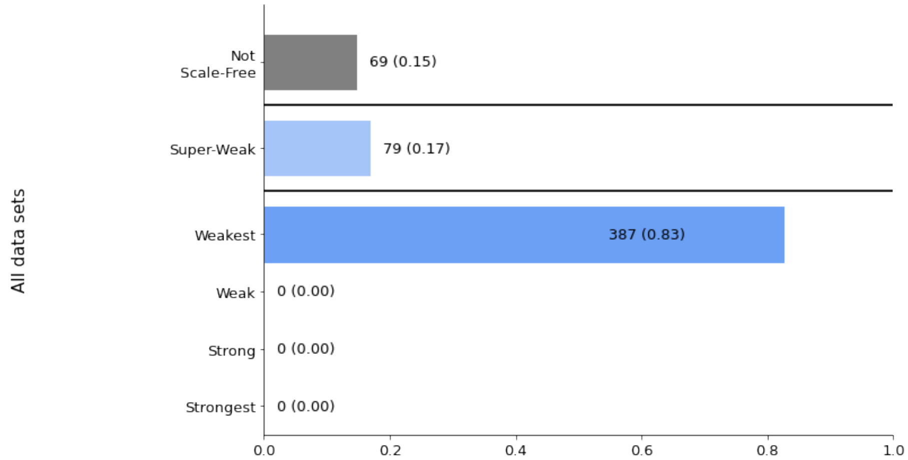


Figure 3.9: Scale-free networks degrees results for cumulative monthly networks

Figure 3.9 shows the results for the monthly cumulative networks. In the system there are no strongest, strong and weak scale-free patterns. For the majority, 387 networks are weakest scale-free, 79 super-weak scale-free. In addition, there are 69 networks are not scale-free.

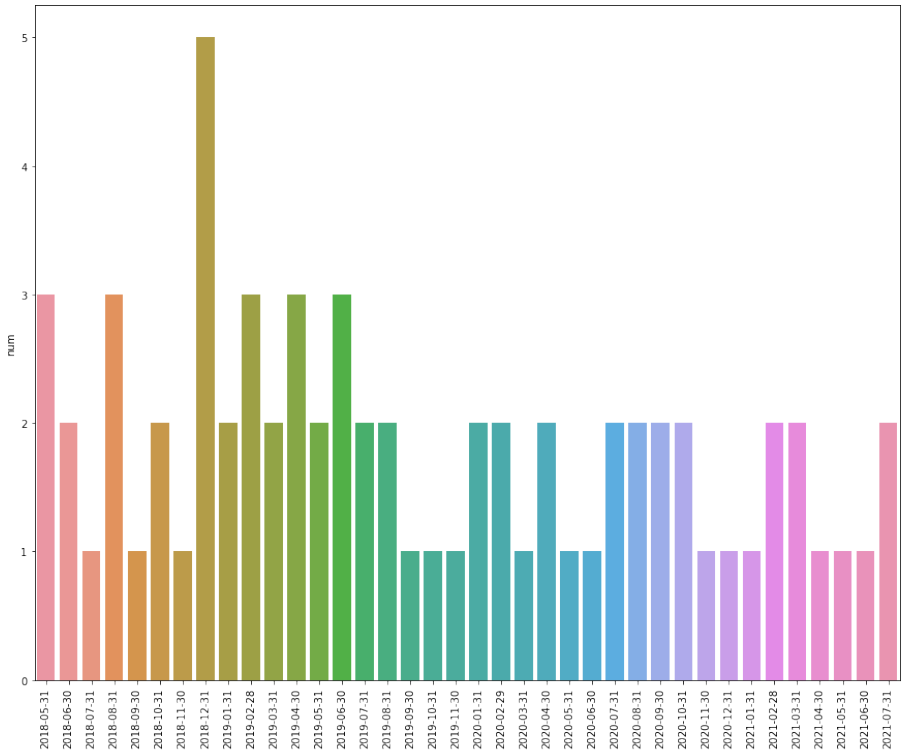


Figure 3.10: Non-scale-free distribution over time for cumulative monthly networks

Figure 3.10 shows the distribution of non-scale-free networks for monthly cumulative graphs over time. Networks, as time passes, become larger and exhibit more scale-free behavior.

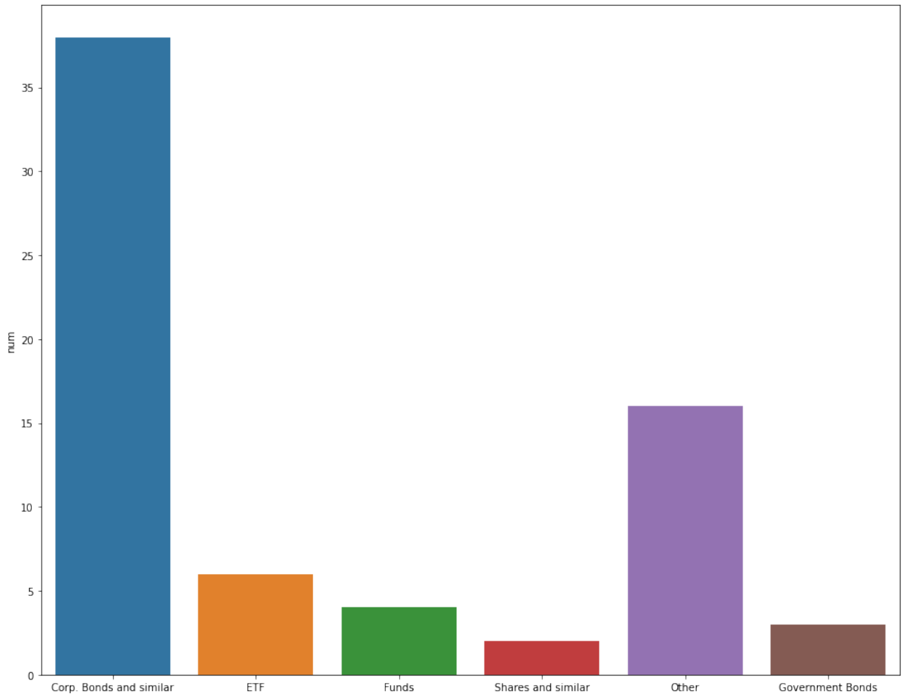


Figure 3.11: Non-scale-free distribution divided by financial instrument type for cumulative monthly networks

Figure 3.11 shows the distribution of non-scale-free networks for monthly cumulative graphs with detail on financial instrument type. Scale-free behavior is not present mostly in corporate bonds and other networks, while shares, government bonds, funds and ETF have more scale free behavior.

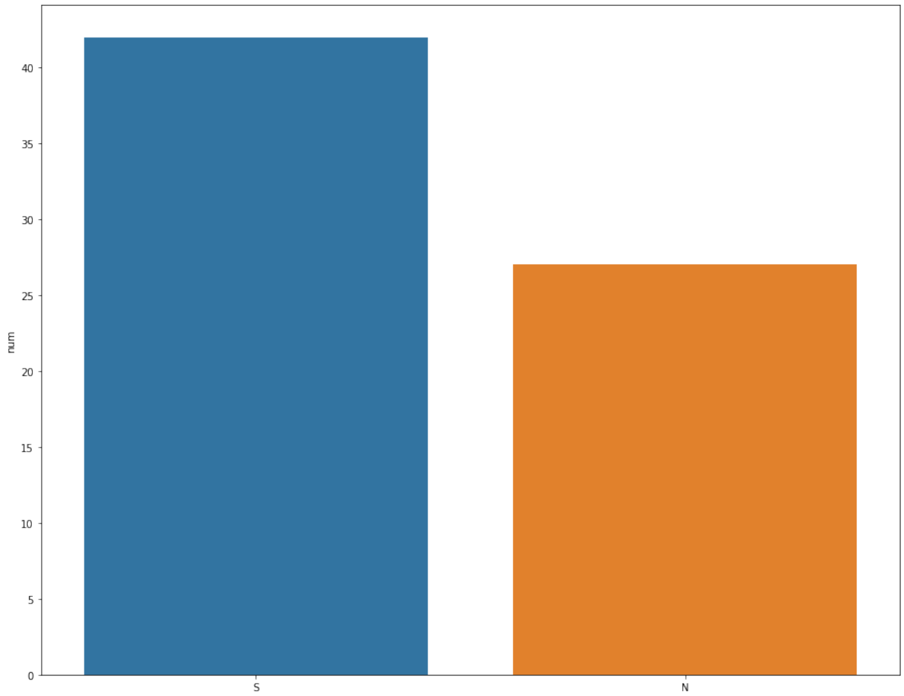


Figure 3.12: Non-scale-free distribution divided by settled and failed instructions for cumulative monthly networks

Figure 3.12 shows non-scale-free networks distribution with a focus on only-failed and only-settled networks. Networks constructed from settled instruction are more non-scale-free compared to those constructed with failed instructions.

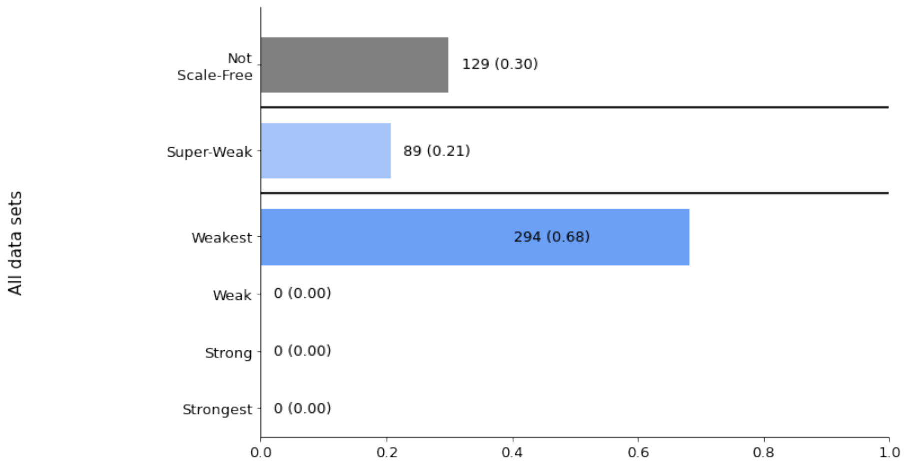


Figure 3.13: Scale-free networks degrees results for non-cumulative monthly networks

Figure 3.13 shows the results for the monthly cumulative networks. There are no strongest, strong and weak scale-free pattern. 294 networks are weakest scale-free, 89 super-Weak scale-free. Moreover, there are 129 Networks are not scale-free.

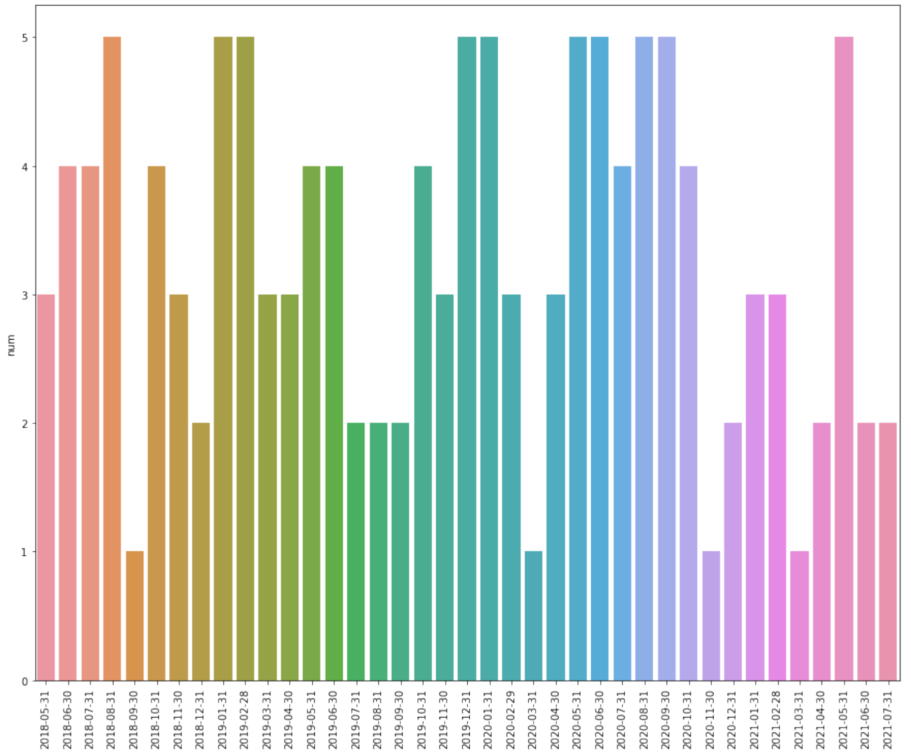


Figure 3.14: Non-scale-free distribution over time for non-cumulative monthly networks

Figure 3.14 shows the distribution of non-scale-free networks for monthly cumulative graphs over time. No obvious pattern emerges.

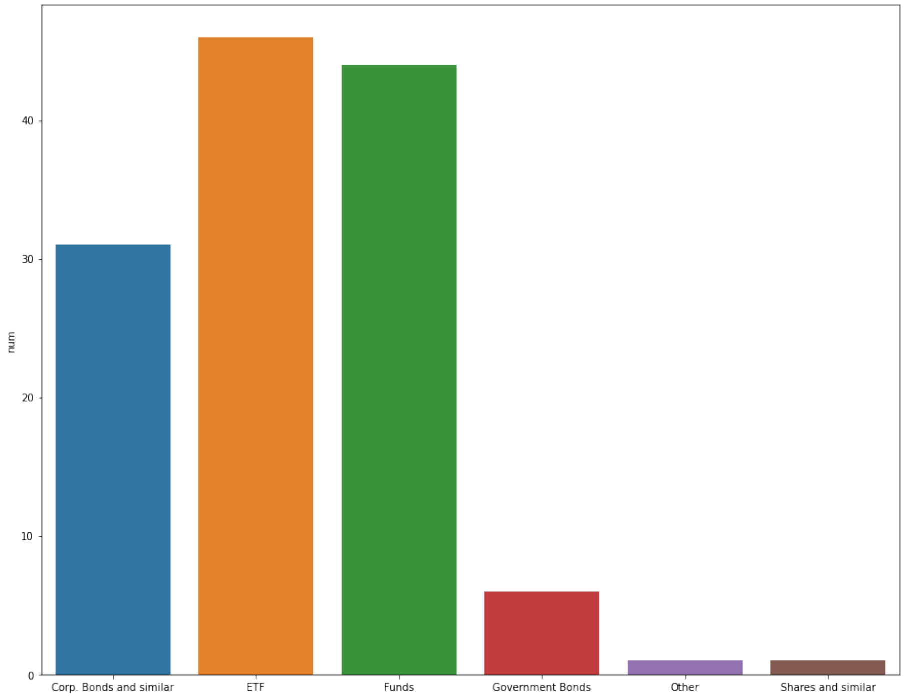


Figure 3.15: Non-scale-free distribution divided by financial instrument type for non-cumulative monthly networks

Figure 3.15 shows the distribution of non-scale-free networks for monthly graphs with a focus on the different financial instrument. Scale free behavior is not present mostly in corporate bonds, ETF and funds, while shares, government bonds, funds and others mostly have a scale free behavior.

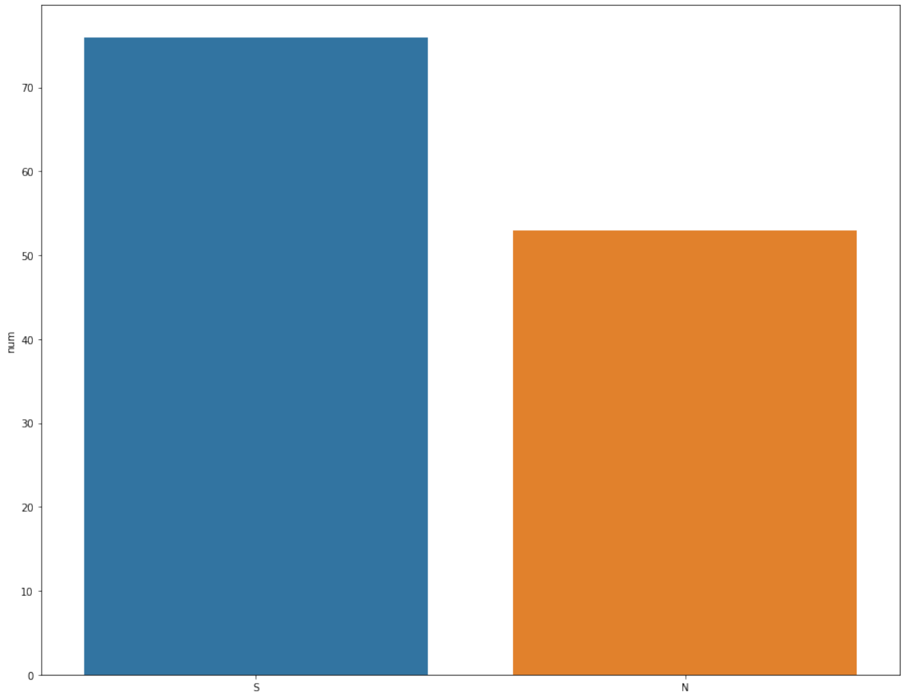


Figure 3.16: Non-scale-free distribution divided by settled and failed instructions for non-cumulative monthly networks

Figure 3.16 shows non-scale-free networks distribution with a focus on settled and failed instructions. Networks constructed with only-settled instructions exhibit a higher number of non-scale-free networks than those constructed with only-failed.

Between monthly cumulative and non-cumulative networks, we identify several common behaviors and unevenness. Monthly cumulative networks in general exhibit a greater number of weakest and super-weak scale-free behaviors, while non-cumulative networks show a greater number of non-scale-free networks. For monthly cumulative non-scale-free networks, we identify a pattern as time pass: as the nodes and links increase, the number of scale-free networks increases. For non-cumulative, instead, no evident patter is detected over time. Moreover, focusing on financial instrument type, we identify similarities such as less scale-free behavior for corporate bonds, while disparities for ETF and funds. For the cumulative and non-cumulative networks we identify similar behaviors for only-settled and only-failed instructions. In summary, both cumulative and non-cumulative networks result with a scale-free distribution. In particular, for non-cumulative almost 60% are weakest, while for cumulative more than 70% are weakest.

3.3.4 Networks Resiliency

One of the characteristics that makes a hub or a key node important is a high betweenness centrality, not only a high degree. Hubs often connect clusters of sub-areas of the graph that are not directly connected to each other. These central nodes are important because they shorten path lengths by making high reachability and fast movement of information. They can be important as brokers and key-players that connect the graph because of their betweenness.

We investigate if the structure of the network become divided into disconnected clusters after a node removal. As mentioned in Section 3.3.2, there exist several approaches to find key nodes in the network. Central nodes, may act as enablers between groups that would otherwise be disconnected. We found these nodes that represent a vulnerable part of the network. In a network composed by financial actors, the deletion of a node in the worst case means that a company has gone bankrupt or, more commonly, the company decides to exit the system. This deletion may alter the structure of the networks. Detecting changes in the connected component (CC) of a network can suggest the presence of structure alterations. After deletion of a node, if there is a change in the connected component, it means that the network is compromised and, moreover, this node is identified as vulnerable. Conversely, if after a deletion, there are no changes in the connected component, this implies that the networks have not been compromised. In addition, the scale-free property is strongly correlated with network robustness to failure[17]. In particular, scale-free networks are more vulnerable to localized attacks, while they are more robust to random failures. The resilience of the network to failures or attacks can then be analyzed by studying how the size of networks components varies as a function of the number of removed nodes.

Different node deletion approaches are applied:

- Random Node deletion: a node is deleted randomly from the Network.
- Localized Node deletion: a deletion of a precisely selected node.

Localized Nodes deletion

Localized Nodes deletion implies a deletion of a precisely selected node. In this case, the selected nodes are central nodes that may alter the network structure after deletion. In particular, given a weakly connected network, it could be considered damaged if after deletion, the weakly connected graph presents any isolated nodes.

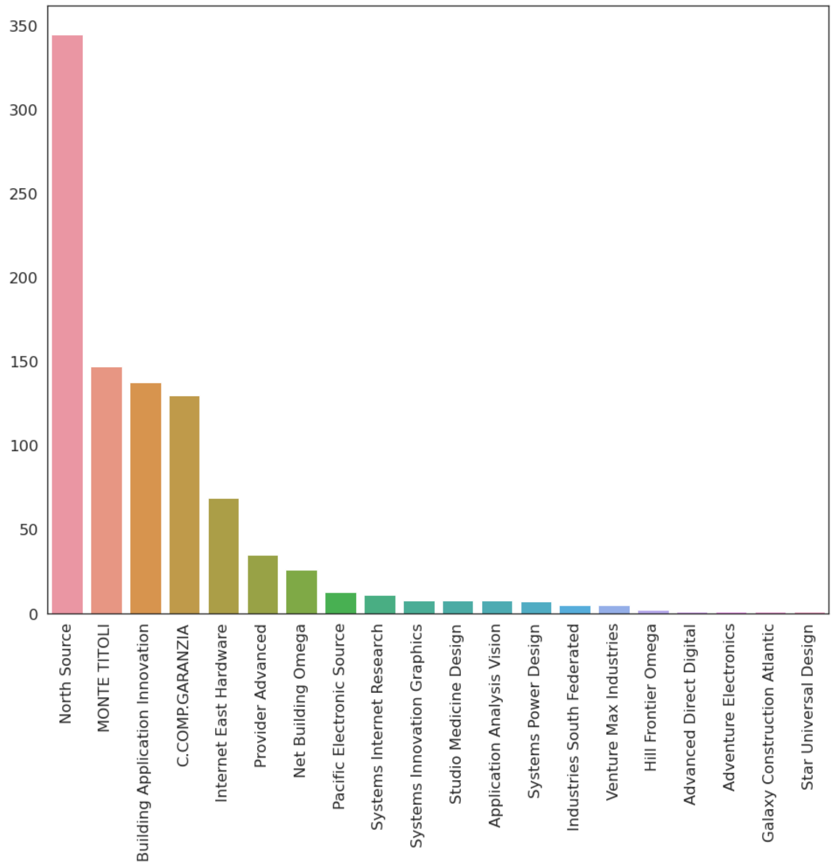


Figure 3.17: Most vulnerable central nodes in cumulative monthly networks

Figure 3.17 shows the frequency of damage in networks when a specific company is deleted. For cumulative networks, North Source, Monte Titoli, and Building Application Innovation are the central nodes that in most cases, if deleted, networks result with isolated nodes.

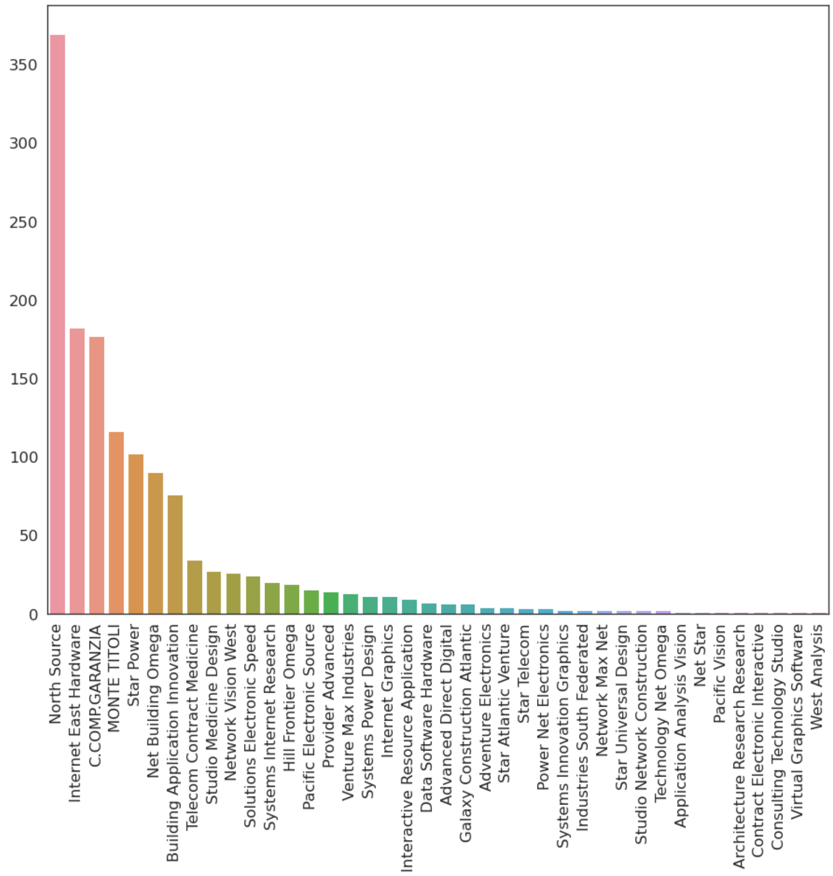


Figure 3.18: Most vulnerable central nodes in non-cumulative monthly networks

From Figure 3.18 it is possible to see the frequency of damage in networks when a specific company is deleted. For non-cumulative networks, North Source, Internet East Hardware and CC&G are the central nodes that in most cases, if deleted, networks result with isolated nodes.

Topology changes after deletion

An in-depth analysis is performed on topology changes after node removal, considering both random and central nodes removal. Two topology measure are examined: size and average shortest path length. At each step 5% of total nodes are eliminated and the corresponding network topology metric is computed.

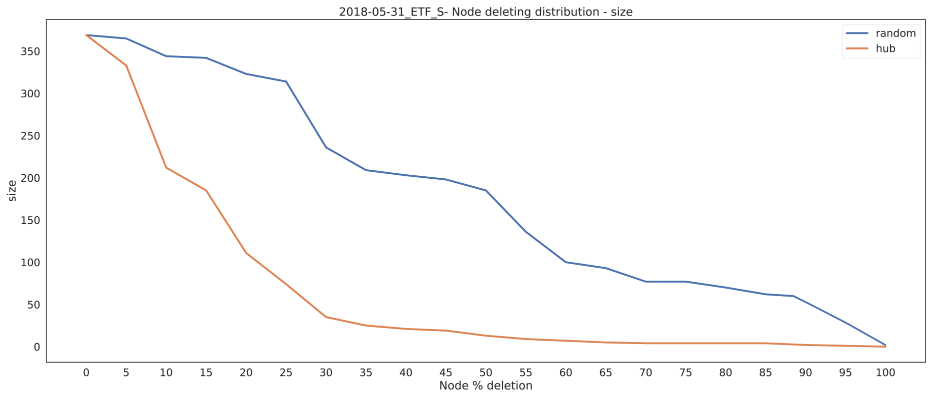


Figure 3.19: Size distribution for ETF network on May 31^{st} 2018 with only-settled upon removal of a certain percentage of nodes

Considering the monthly cumulative data on May 31^{st} 2018, Figure 3.19 shows the size distribution for ETF network with only-settled instructions as the percentage of node elimination increases. Hub deletion quickly reduces the network size. After deleting 30% of the nodes, the size decreases to less than 50 links. Random node deletion results in a slower decrease in network size. More than 100 links remain even after random removal of more than 60% of the nodes.

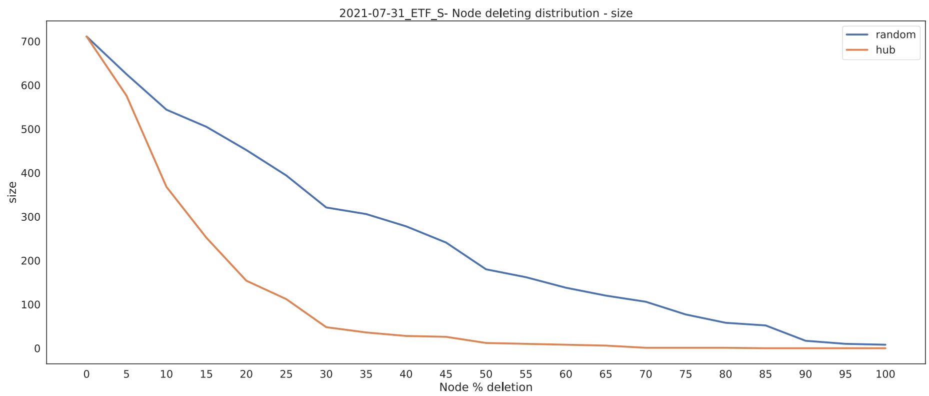


Figure 3.20: Size distribution for ETF network on July 31^{st} 2021 with only-settled upon removal of a certain percentage of nodes

Considering the monthly cumulative data on July 31^{st} 2021, Figure 3.20 shows the distribution of ETF network size with only-settled instructions as the percentage of node elimination increases. The results are similar to Figure 3.20, however the curve is smoother as the number of nodes and links is greater since the network is generated from two cumulative years of data.

The average shortest path length is one of the three most relevant measures of network topology, along with the clustering coefficient and degree distribution. It is defined as the sum of path lengths between all pairs of nodes normalized by $n \times (n-1)$, where n is the number of nodes. Moreover, it represents the ability of two nodes to communicate with each other: the smaller the distance is, the shorter is the expected path between them. Networks with a very large number of nodes can have fairly a small diameter[37]. Referring to Section 2.2, they used Average shortest path length to evaluate how networks topology changes after node deletion.

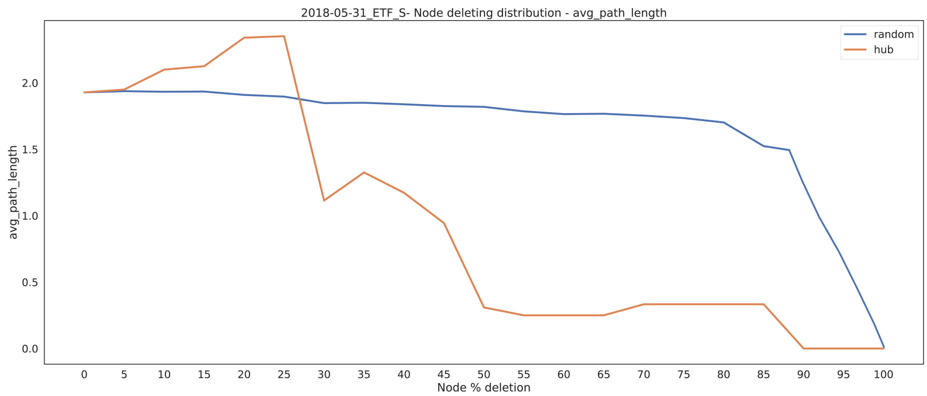


Figure 3.21: Average shortest path length distribution for ETF network on May 31^{st} 2018 with only-settled upon removal of a certain percentage of nodes

Figure 3.21 shows how the average shortest path length changes deleting random and central nodes on monthly cumulative ETF network with only-settle instruction on May 31^{st} 2018. The elimination of hub nodes results in increased distance in the network. The reason is that by removing the central node links, fewer paths are available to connect peripheral nodes. After reaching about 25% of node removal, the average shortest path length is significantly reduced. Random removal, on the other hand, presents a stable shortest path length. The same behavior is encountered in the article "On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia" and shown in Figure 2.1.

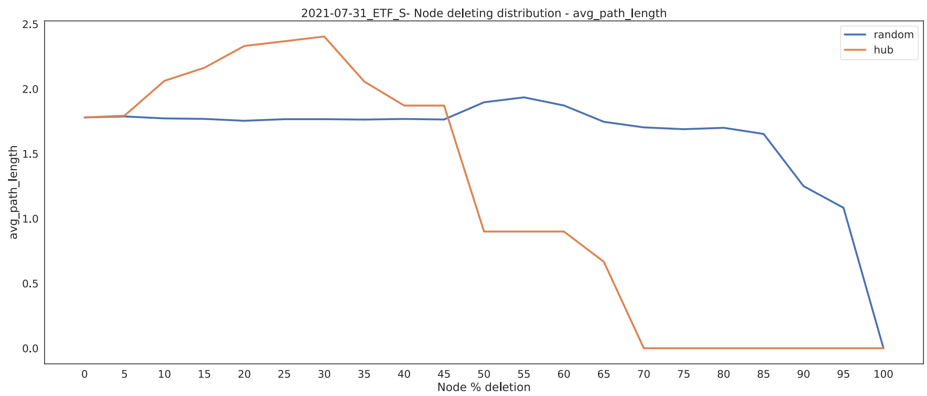


Figure 3.22: Average shortest path length distribution for ETF network on July 31^{st} 2021 with only-settled upon removal of a certain percentage of nodes

Figure 3.22 shows how the average shortest path length changes deleting random and central nodes on monthly cumulative ETF network with only-settle instruction in July 31^{st} 2021. In July, it is observed similar behavior to the May 31^{st} 2018 network, however the increase of shortest path length persists up to the 30% of node deletion and decreases rapidly after the 45% cut-off is reached. After 70% of hubs deletion, the average shortest path length zeroes out.

In conclusion, for monthly graphs we identify the most important topological measures with no distinction between monthly cumulative and non-cumulative. In particular, the number of companies and trades, but also measure such as clustering coefficient, assortativity, transitivity and connected components. Distinguishing between cumulative and non-cumulative networks the centrality analysis shows that for cumulative monthly networks, South Construction, Net Venture, Research System Design are the most frequent central nodes, while for non-cumulative South Construction, Net Venture and Cassa Compensazione e Garanzia are the most frequent central nodes. CC&G has a high centrality because is the Italian entity deputed to clearing and netting. Analysing the centrality over time, for ETF non-cumulative networks we can see a non-stable pattern of firms' positions over time. For ETF cumulative, instead, companies ranking is quite stable. Moreover, Monthly cumulative networks in general exhibit a greater number of weakest and super-weak scale-free behaviors, while non-cumulative networks show a greater number of non-scale-free networks. For cumulative networks, North Source, Monte Titoli, and Building Application Innovation are the most central nodes that in most cases, if deleted, networks result with isolated nodes, while for non-cumulative are North Source, Internet East Hardware and CC&G. We identify similar behavior of "On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia" regarding networks size and average shortest path length distributions after node elimination. In general, considered networks are more vulnerable to localized attacks on central nodes, while they are more robust to random failures.

3.4 Daily Graph

Two case studies are performed for the daily graph compared with the monthly graph, specifically with a focus on disruptive events. Disruptive events could irreversibly change the topology and structure of the networks. Referring to the article *"The Topology of Interbank Payment Flows"* in Section 2.3, events such as the September 11^{th} 2001 terrorist attack demonstrated how these phenomena can alter economics systems. In this case the financial systems of the USA were disrupted.
Two different case-studies of disruptive events:

- The impact of Covid-19 on network metrics and topology during the first lockdown: Period from January 2020 to June 2020. During this period, several topological measure are analyzed such as mean degree, assortativity, clustering coefficient and transitivity. In particular, the average trend over time of daily networks with and without distinction on financial instrument type and settlement status.

- The impact of a large emission of BTP Italia and BTP Futura: the analysis considers consideration five days before and after the announcement of the BTP. This is considered a disruptive event since large amounts of instructions are exchanged during the emission dates. For each period of emission, the mean degree of government bond networks is analyzed before and after BTP release with the distinction between only-failed and only-settled instructions. Only government bond networks are considered because BTPs are included these financial instrument category.

3.4.1 Case study: Covid-19

The Covid-19 breakout challenges all areas of economics, including health, industrial organization, macroeconomics, finance, history, development, inequality, political economy and public finance, and concerns theory as well as empirical evidence. The spread of the virus fostered social estrangement which led to the shutdown of financial markets, businesses and events [38].

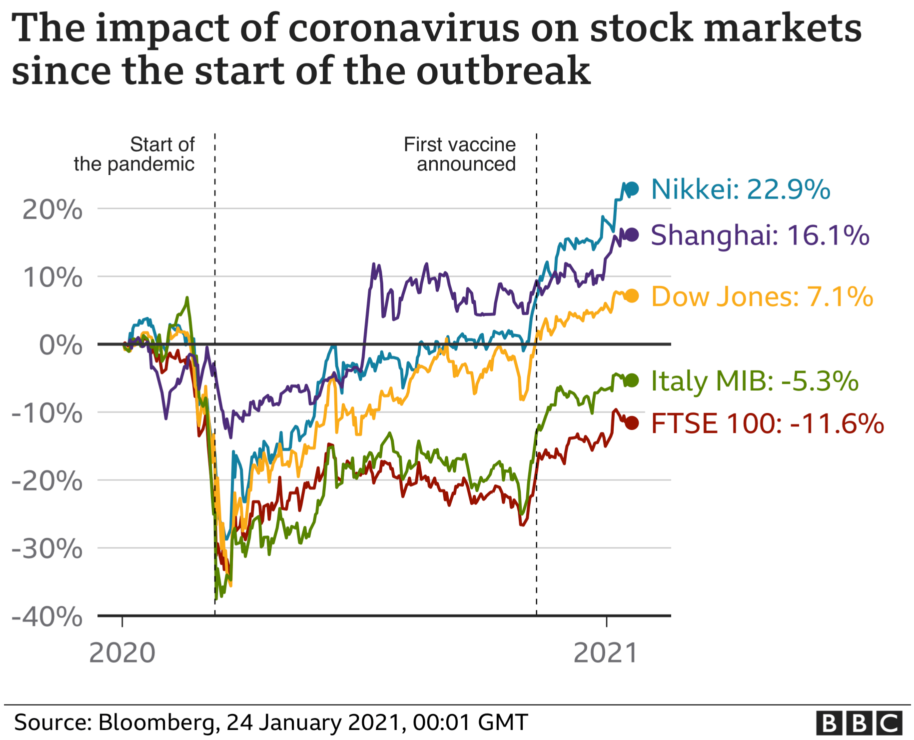


Figure 3.23: The impact of coronavirus on stock markets by Bloomberg

Figure 3.23 shows the Covid-19 impact on stock markets [39]. Shifts caused by Covid-19 in stock markets, where companies' shares are bought and sold, can affect the value of pensions or individual savings accounts. The FTSE, Dow Jones Industrial Average and the Nikkei experienced large declines due to the increased number of Covid-19 cases in the early months of the crisis. Major Asian and U.S. stock markets rallied after the announcement of the first vaccine in November. The FTSE fell 14.3% in 2020, its worst performance since 2008.

The objective of the case-study is to evaluate the impact of Covid-19 on the network metrics and topology during the first lockdown and thereafter. The time interval for constructing daily networks is from January 1^{st} 2019 to December 31^{st} 2020. We considered February 1^{st} 2020 as the Italian starting date for Covid-19 and we compared the connections and the size of the networks over time.

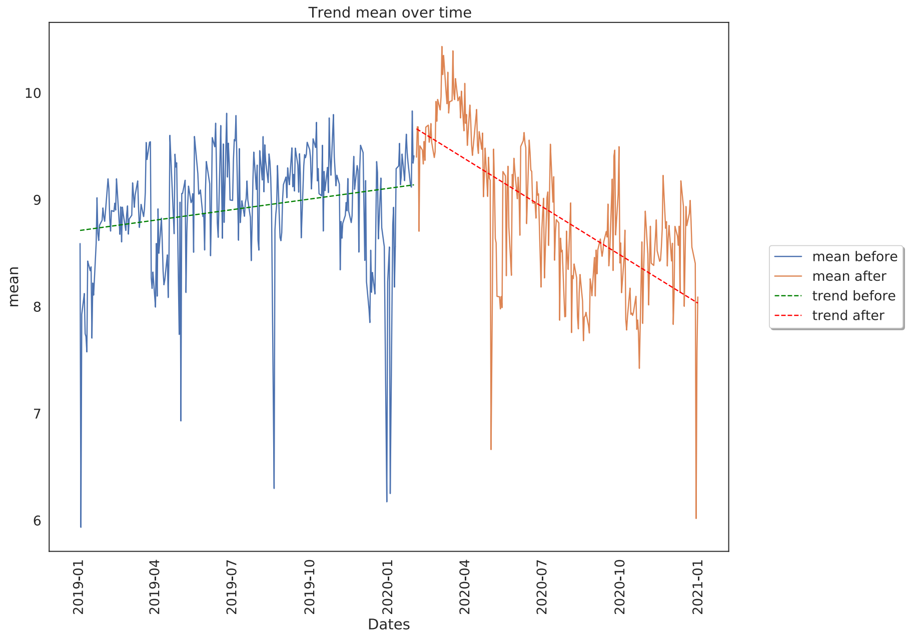


Figure 3.24: Daily networks overall mean degree trend over time before and during coronavirus

Figure 3.24 shows the daily networks overall mean degree trend over time, without distinguishing between financial instruments and settlement status of the instruction. After February 1^{st} 2020, a downward trend in the overall mean degree is observed. The negative peaks represent particular business days such as New Year's Eve, Easter and summer holidays.

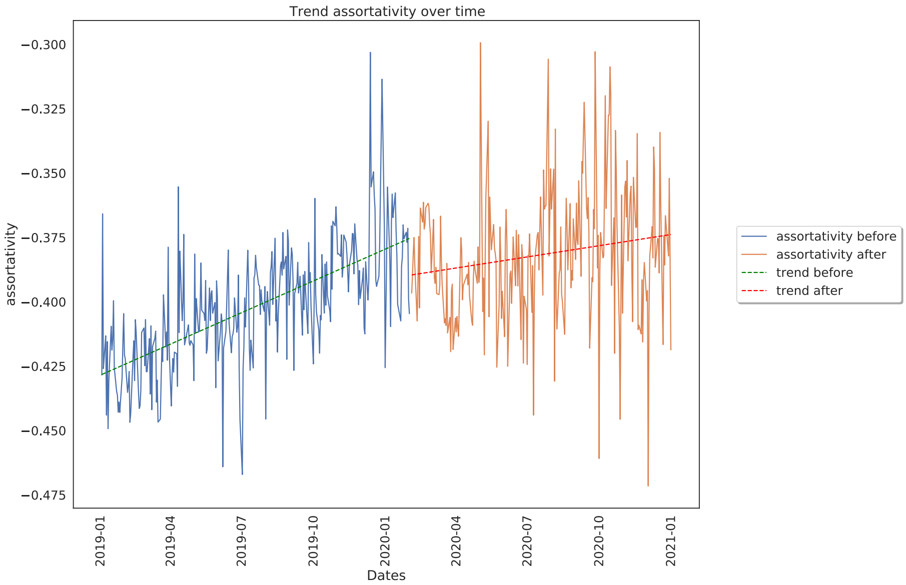


Figure 3.25: Daily networks overall assortativity trend over time before and during coronavirus

Figure 3.25 shows the assortativity for daily aggregated networks without distinction between settled/failed instructions and type of financial instrument. In the illustration, assortativity before Covid-19 is increasing, while during the pandemic it slowly increases. In the 2019 assortativity in the networks had an increasing trend, while during coronavirus the trend decreased. This mean that companies in 2019 slightly reduce the tendency of trading with companies with an extremely different degree.

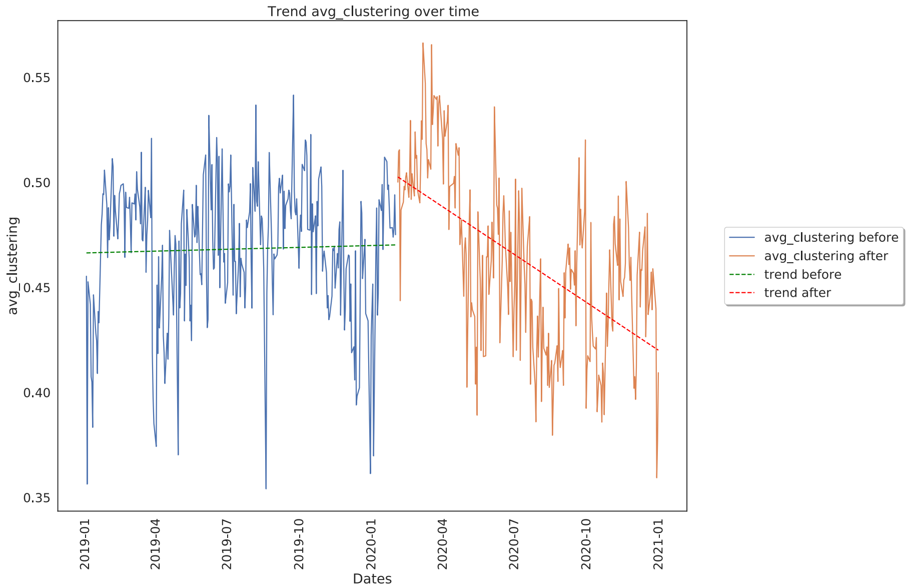


Figure 3.26: Daily networks overall average clustering coefficient trend over time before and during coronavirus

Figure 3.26 shows the average clustering coefficient for daily aggregated networks without distinction between settlement status of the instructions and type of financial instrument. Before Covid-19 the trend is stable, while during the pandemic the clustering coefficient trend is decreasing. Companies cluster to a lesser extent during coronavirus than in previous periods.

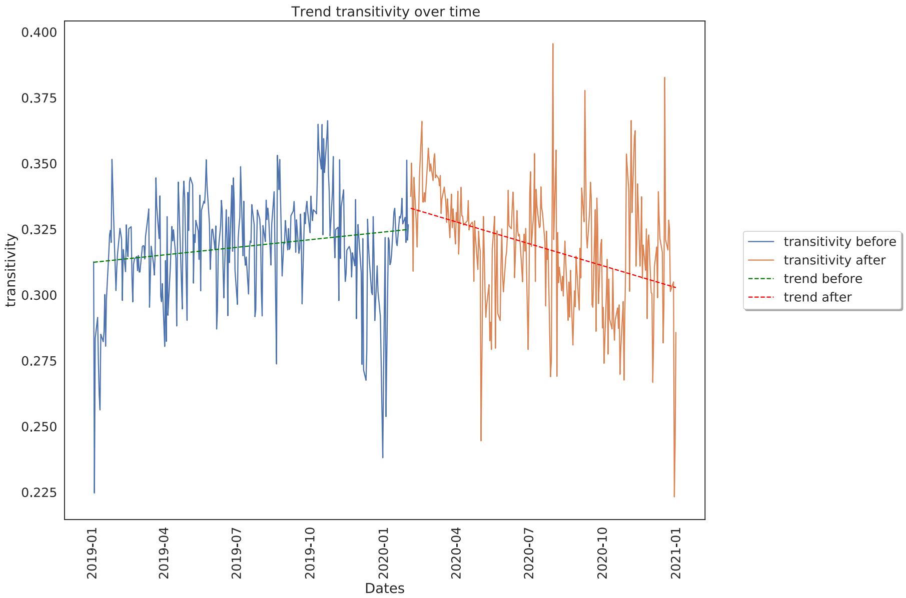


Figure 3.27: Daily networks overall transitivity trend over time before and during coronavirus

Figure 3.27 shows the transitivity for daily aggregated networks without distinction between settled/failed instructions and type of financial instrument. Before Covid-19 the trend is stable and slightly increasing, while during the pandemic the transitivity is decreasing.

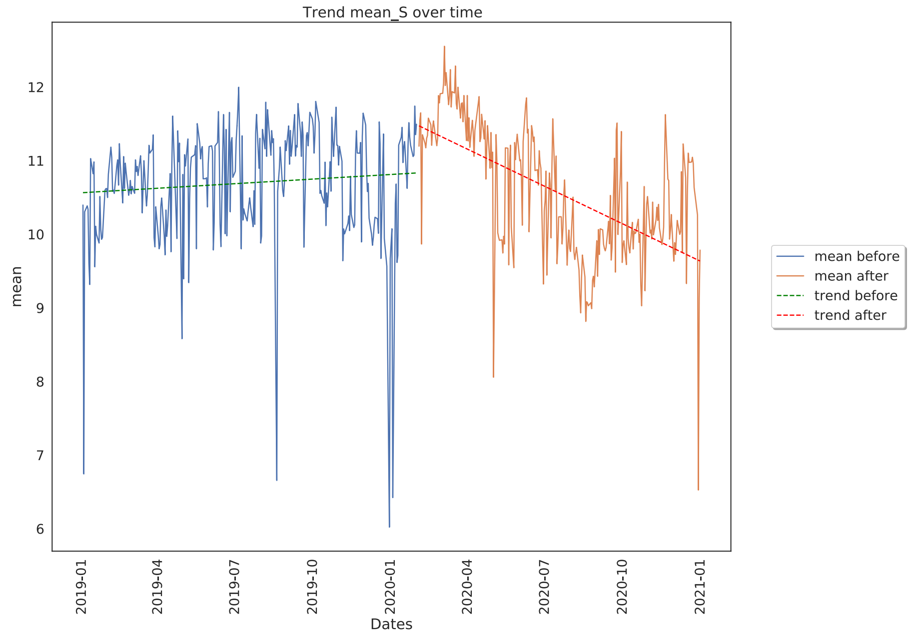


Figure 3.28: Only-settled daily networks overall mean degree trend over time before and during coronavirus

Figure 3.28 shows the mean degree for daily aggregated networks for settled instructions without distinction by type of financial instrument. Before Covid-19 the trend is sightly increasing, while during the pandemic the mean degree trend is decreasing.

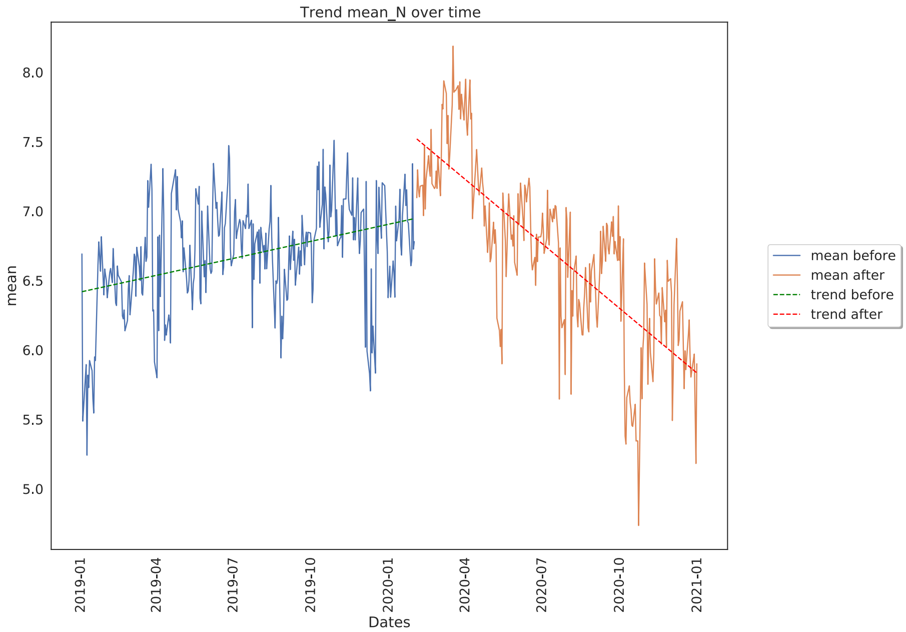


Figure 3.29: Only-failed daily networks overall mean degree trend over time before and during coronavirus

Figure 3.29 exhibit the mean degree for daily aggregated networks with only-failed instructions and without distinction between type of financial instrument. Before Covid-19 the trend is increasing, while during Covid-19 the mean degree trend is decreasing.

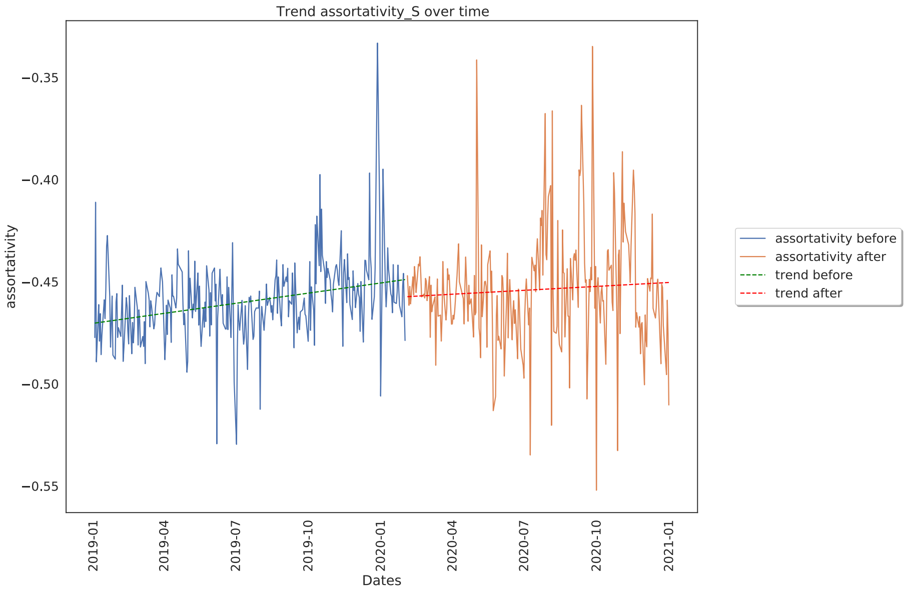


Figure 3.30: Only-settled daily networks overall assortativity trend over time before and during coronavirus

Figure 3.30 shows the assortativity for daily aggregated networks with only-settled instructions and without distinction between type of financial instrument. Assortativity before and during Covid-19 is stable, increasing slightly. The assortativity is not much different for only-settle instructions.

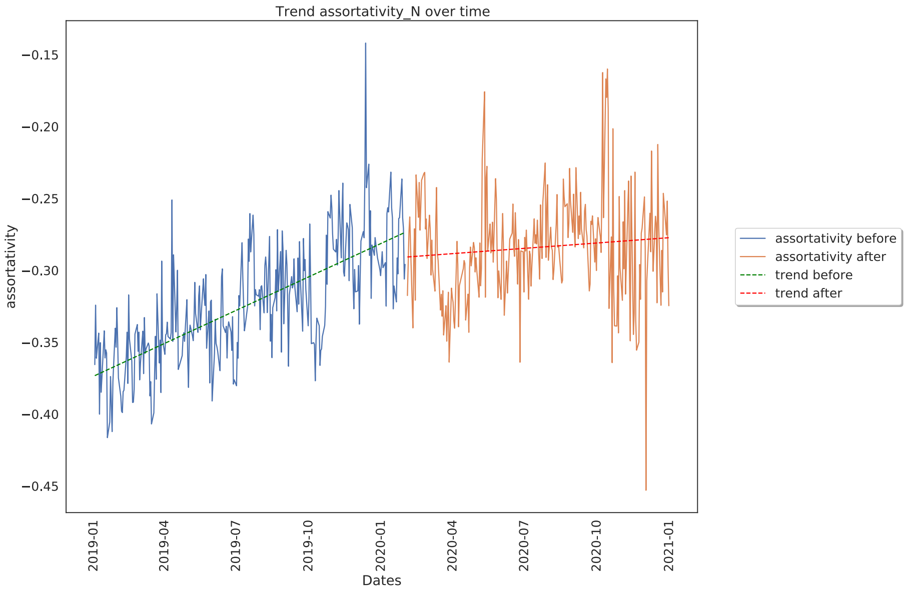


Figure 3.31: Only-failed daily networks overall assortativity trend over time before and during coronavirus

Figure 3.31 shows the assortativity for daily aggregated networks for failed instructions without distinction for type of financial instrument. Assortativity before and after Covid-19 is increasing, while after Covid-19 is stable and slightly increasing.

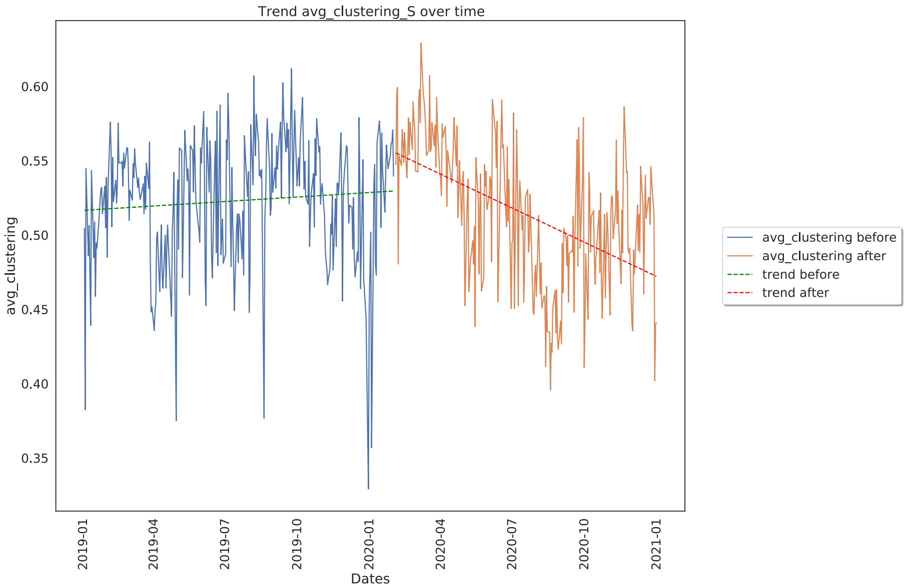


Figure 3.32: Only-settled daily networks overall average clustering coefficient trend over time before and during coronavirus

Figure 3.32 shows the average clustering coefficient for daily aggregated networks with only-settled instructions and without distinction between type of financial instrument. Before Covid-19 the trend is stable, while during the pandemic the clustering coefficient trend is decreasing.

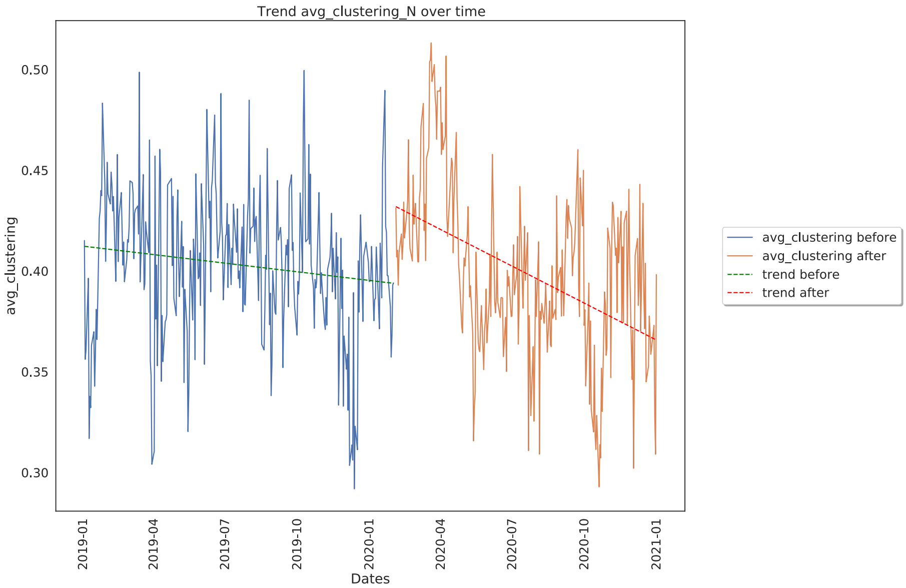


Figure 3.33: Only-failed daily networks overall average clustering coefficient trend over time before and during coronavirus

Figure 3.33 shows the average clustering coefficient for daily aggregated networks with only-failed instructions and without distinction between type of financial instrument. Before and during Covid-19 the clustering coefficient trend is decreasing.

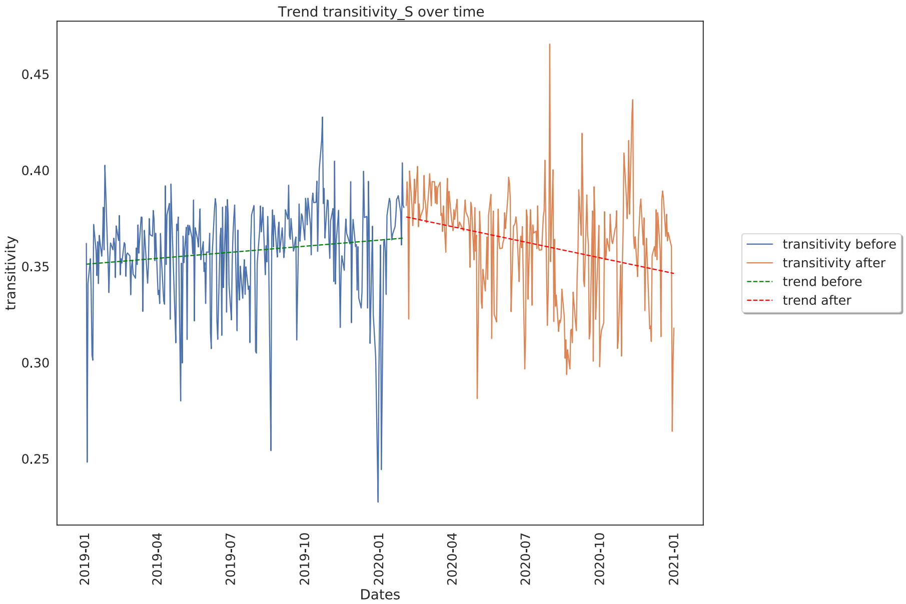


Figure 3.34: Only-settled daily networks overall transitivity trend over time before and during coronavirus

Figure 3.34 shows the transitivity for daily aggregated networks with only-settled instructions and without distinction between type of financial instrument. Before Covid-19 the trend results stable, slightly increasing while during Covid-19 the transitivity trend is decreasing.

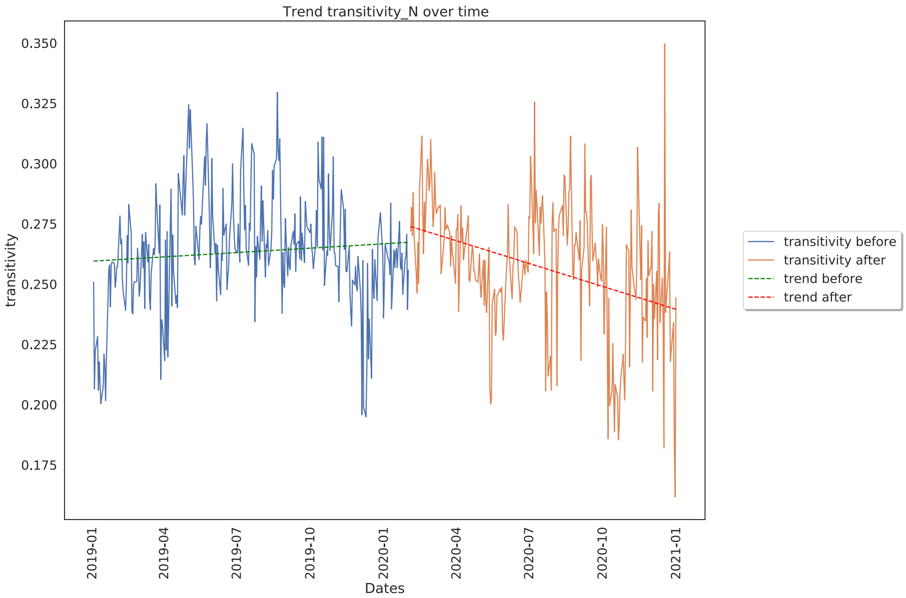


Figure 3.35: Only-failed daily networks overall transitivity trend over time before and during coronavirus

Figure 3.35 shows the transitivity for daily aggregated networks with only-settled instructions and without distinction between type of financial instrument. Before Covid-19 the trend is stable, slightly increasing while during the pandemic the trend of the transitivity is decreasing.

	Mean Degree			Assortativity		
	before	during	Δ	before	during	Δ
ETF_N	9.803	9.629	-0.173	-0.428	-0.411	0.017
ETF_S	15.974	15.284	-0.69	-0.502	-0.507	-0.005
$\text{Government Bonds}_S$	22.217	22.138	-0.08	-0.555	-0.554	0.001
$\text{Government Bonds}_N$	10.632	10.383	-0.249	-0.463	-0.402	0.061
Shares_S	22.453	21.319	-1.134	-0.472	-0.49	-0.018
Shares_N	12.711	12.843	0.131	-0.447	-0.435	0.013
Corp. Bonds_S	8.098	7.784	-0.315	-0.61	-0.602	0.008
Corp. Bonds_N	7.619	6.745	-0.874	-0.345	-0.291	0.054
Funds_S	15.579	14.811	-0.768	-0.491	-0.491	0.0
Funds_N	9.948	9.803	-0.145	-0.438	-0.412	0.026
Other_S	4.515	6.511	1.996	-0.537	-0.479	0.058
Other_N	3.134	2.987	-0.147	-0.184	-0.129	0.055

Table 3.6: Average degree and assortativity before and during coronavirus and the difference

	Average Clustering			Transitivity		
	before	during	Δ	before	during	Δ
ETF_N	0.594	0.588	-0.006	0.264	0.275	0.011
ETF_S	0.703	0.665	-0.037	0.407	0.424	0.018
Government $Bonds_S$	0.673	0.671	-0.003	0.439	0.443	0.004
Government $Bonds_N$	0.605	0.575	-0.03	0.297	0.295	-0.002
$Shares_S$	0.771	0.754	-0.016	0.568	0.559	-0.009
$Shares_N$	0.607	0.588	-0.019	0.343	0.353	0.01
Corp. $Bonds_S$	0.274	0.261	-0.013	0.261	0.254	-0.007
Corp. $Bonds_N$	0.484	0.432	-0.052	0.306	0.269	-0.036
$Funds_S$	0.673	0.633	-0.039	0.402	0.416	0.014
$Funds_N$	0.613	0.601	-0.012	0.259	0.274	0.015
$Other_S$	0.188	0.294	0.105	0.217	0.325	0.108
$Other_N$	0.105	0.065	-0.039	0.145	0.054	-0.091

Table 3.7: Mean average clustering and transitivity before and during coronavirus and the difference

As shown in Table 3.6 coronavirus pandemic impacts Monte Titoli's networks, with a visible decrease in mean degree and an increase in assortativity. This means networks are less connected and less disassortative. Only-settled other networks is the only network that in which mean degree increases, while only-settled ETF and only-settle shares become slightly more disassortative. As shown in Table 3.7 coronavirus decreases the average clustering coefficient. Only-settled other networks is the only type in which during Covid-19 the average clustering coefficient increases, while no detectable patters are identified for transitivity before and during the pandemic. In general, coronavirus pandemic impacts Monte Titoli's networks, with a decrease in the mean degree, clustering coefficient and disassortativity.

3.4.2 Case study: BTP Italia and BTP Futura emissions

Multi-year Treasury Bonds (Buoni del Tesoro Poliennali - BTP) are medium to long-term debt securities issued by the Italian Department of the Treasury with a fixed deferred coupon paid every six months[40]. It is the first government bond indexed to Italian inflation. It is a bond designed mainly to meet the needs of savers and small investors. BTPs Futura are the only government securities reserved exclusively for individual savers and other comparable investors (retail market). They are called "Futura" (trans. Future) because they are designed to protect savings at the same time to support the "future of the country", with particular reference to overcoming the health and economic crisis caused by COVID-19[41].
The focus of this case-study is to investigate the impact of large government bonds emission, such as BTP Italia and BTP Futura, on the network metrics and topology. The analysis considers the next ten days following the announcement of the BTP. This is considered a disruptive event since a large number of instructions are traded during the emission dates.

BTP Italia

The BTP Italia considered emission dates are:

- From May 14^{th} to May 17^{th} 2018.
- From November 19^{th} to November 22^{nd} 2018.
- From October 21^{st} to October 23^{rd} 2019.
- From May 18^{th} to May 20^{th} 2021.

For each emission date, we consider the state of the networks 5 days before and after the emission. We consider the impact of BTP emission on government bonds networks since BTPs belongs to this financial instrument type.

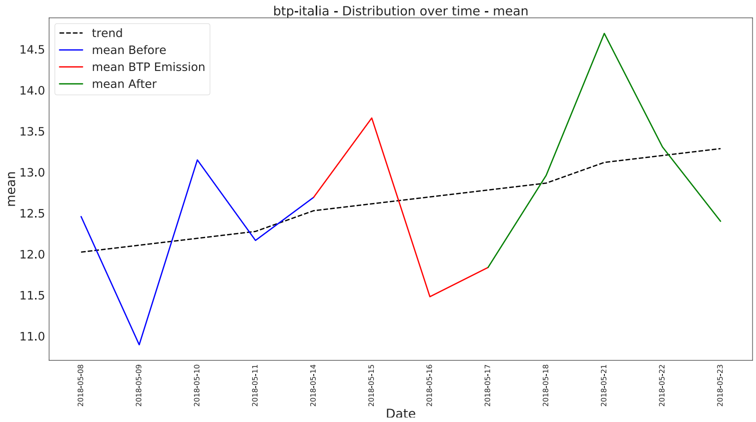


Figure 3.36: Only-settled government bonds network mean degree after May 2018 BTP Italia emissions

Figure 3.36 shows the mean degree distribution for government bonds networks with only-settled instructions during BTP Italia emission dates in May 2018. During emission dates, the network mean degree is increasing, however on May 16^{th} it exhibits a decrease.

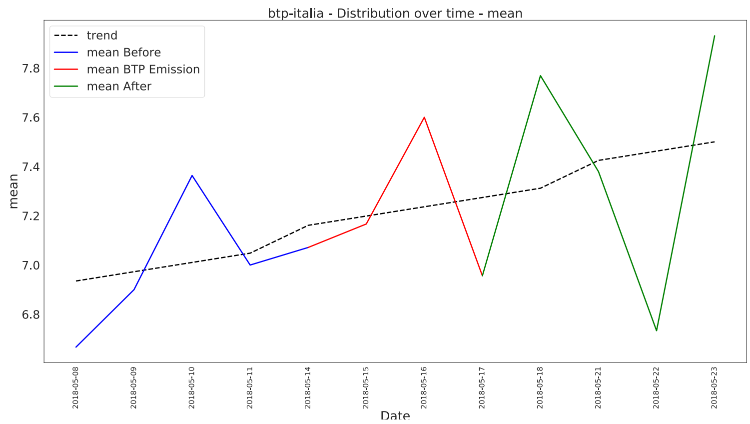


Figure 3.37: Only-failed government bonds network mean degree after May 2018 BTP Italia emissions

Figure 3.37 shows the mean degree distribution for government bonds networks with only-failed instructions during BTP Italia emission dates in May 2018. During emission dates, the network mean degree is increasing, however on May 17^{th} it exhibits a decrease.

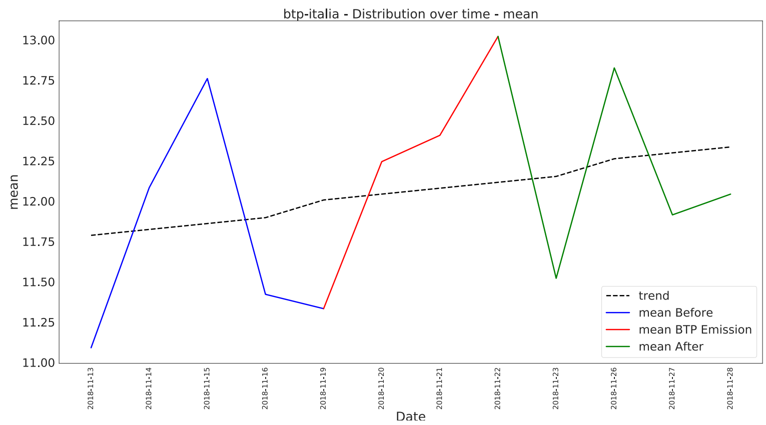


Figure 3.38: Only-settled government bonds network mean degree after November 2018 BTP Italia emissions

Figure 3.38 shows the mean degree distribution for government bonds networks with only-settled instructions during BTP Italia emission dates in November 2018. During emission dates, the network mean degree is increasing, however the following days it decreases.

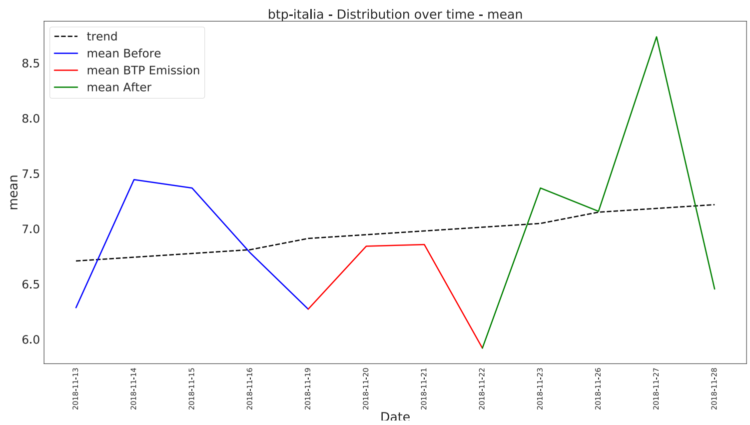


Figure 3.39: Only-failed government bonds network mean degree after November 2018 BTP Italia emissions

Figure 3.39 exhibits the mean degree distribution for government bonds networks with only-failed instructions during BTP Italia emission dates in November 2018. During emission dates, the network mean degree is increasing and stable on November 21^{st}, however on November 22 th it decreases.

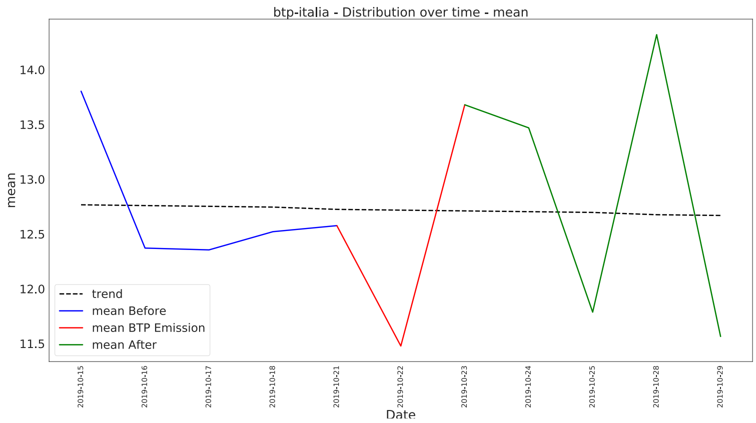


Figure 3.40: Only-settled government bonds network mean degree after October 2019 BTP Italia emissions

Figure 3.40 shows the mean degree distribution for government bonds networks with only-settled instructions during BTP Italia emission dates in October 2019. During emission dates, the network mean degree is decreasing, while on November 22^{st} it exhibit an increase.

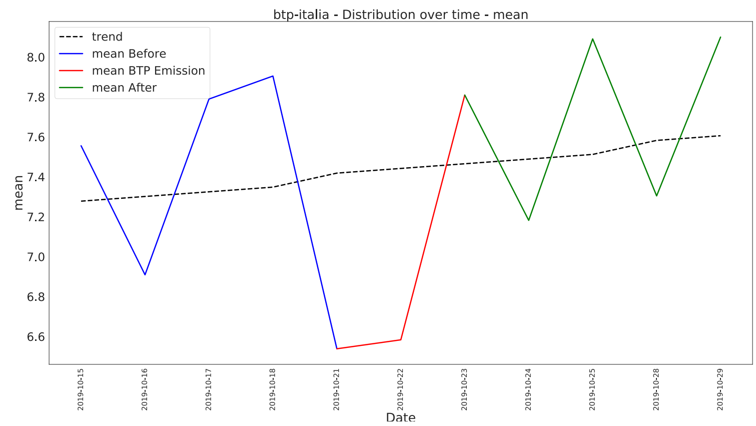


Figure 3.41: Only-failed government bonds network mean degree after October 2019 BTP Italia emissions

Figure 3.41 shows the mean degree distribution for government bonds networks with only-failed instructions during BTP Italia emission dates in October 2019. During emission dates, the network mean degree is increasing, with a spike on October 23^{rd}.

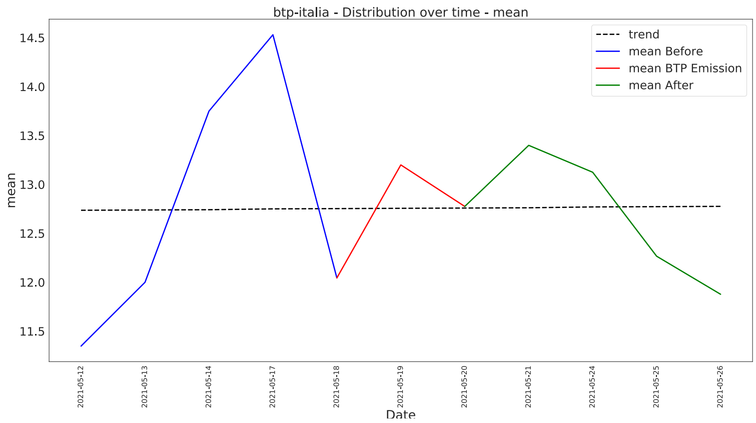


Figure 3.42: Only-settled government bonds network mean degree after May 2021 BTP Italia emissions

Figure 3.42 shows mean degree distribution for government bonds networks with only-settled instructions during BTP Italia emission dates in May 2021. During emission dates, the network mean degree is increasing, while on the last day it slightly decreases.

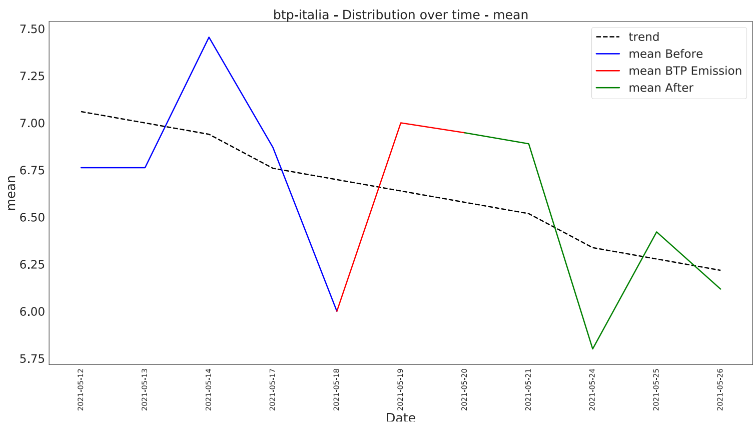


Figure 3.43: Only-failed government bonds network mean degree after May 2021 BTP Italia emissions

Figure 3.43 shows mean degree distribution for government bonds networks with only-failed instructions during BTP Italia emission dates in

May 2021. During emission dates, the network mean degree is increasing, while on the last day it slightly decreases.

BTP Futura

The BTP Futura considered emission dates are:

- From July 6^{th} to July 10^{th} 2020.
- From November 9^{th} to November 13^{th} 2020.
- From April 19^{th} to April 23^{rd} 2021.

For each emission date, we consider the state of the networks 5 days before and after the emission.

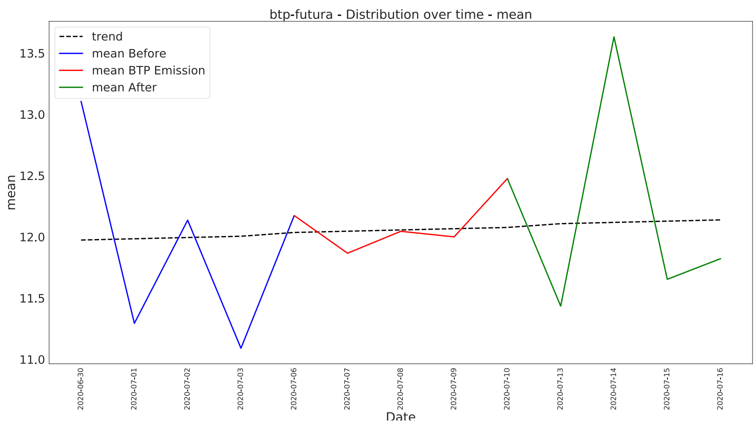


Figure 3.44: Only-settled government bonds network mean degree after July 2020 BTP Futura emissions

Figure 3.44 shows mean degree distribution for government bonds networks with only-settled instructions during BTP Italia emission dates in July 2020. During emission dates, the network mean degree is quite stable, while before and after emission exhibit unsteady behaviors.

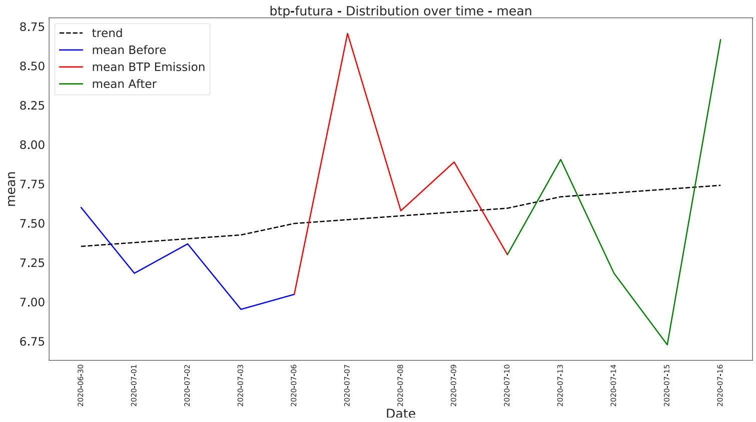


Figure 3.45: Only-failed government bonds network mean degree after July 2020 BTP Futura emissions

Figure 3.45 shows mean degree distribution for government bonds networks with only-failed instructions during BTP Italia emission dates in July 2020. On first day of emission, the network mean degree is largely increasing, while on July 7^{th} it decreases.

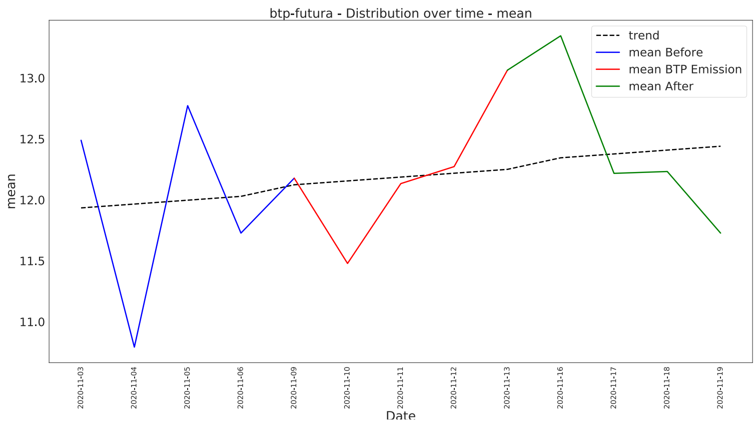


Figure 3.46: Only-settled government bonds network mean degree after November 2020 BTP Futura emissions

Figure 3.46 shows mean degree distribution for government bonds networks with only-settled instructions during BTP Futura emission dates in November 2020. During emission dates, the degree decreases on the first day, however after November 10^{th}, it exhibit an increase.

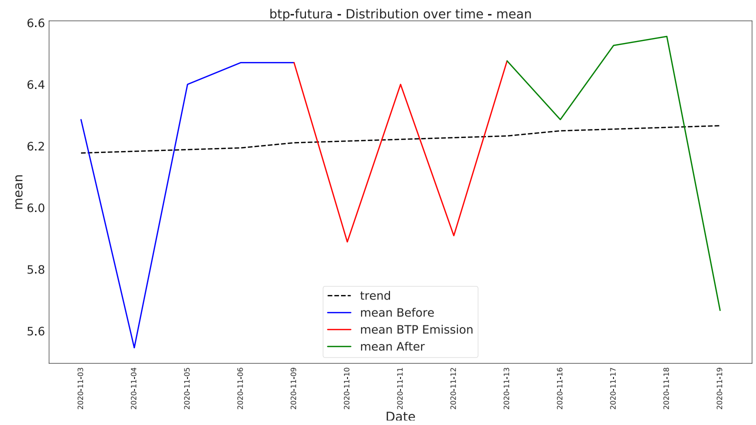


Figure 3.47: Only-failed government bonds network mean degree after November 2020 BTP Futura emissions

Figure 3.47 shows mean degree distribution for government bonds networks with only-failed instructions during BTP Futura emission dates in November 2020. During emission dates, there is a continuous increase and decrease of the mean degree.

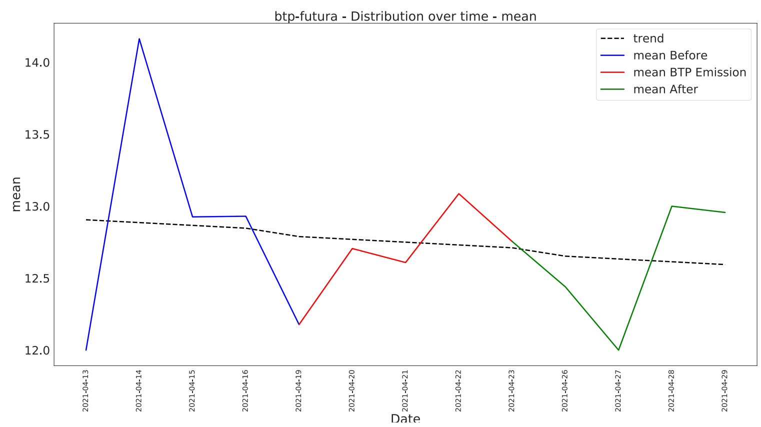


Figure 3.48: Only-settled government bonds network mean degree after April 2021 BTP Futura emissions

Figure 3.48 shows mean degree distribution for Government Bonds networks for settled instructions during BTP Italia emission dates in April 2021. During emission dates, the mean degree is increasing and slightly decreasing after.

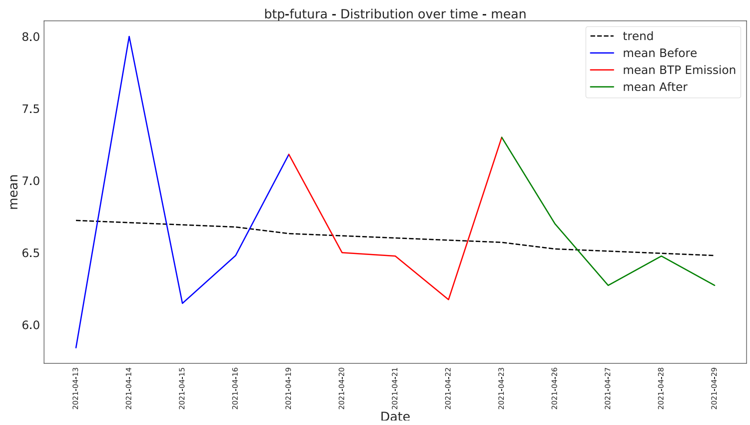


Figure 3.49: Only-failed government bonds network mean degree after April 2021 BTP Futura emissions

Figure 3.49 shows mean degree distribution for government bonds networks with only-failed instructions during BTP Italia emission dates in April 2021. During emission dates, the mean degree is decreasing and in the last day it exhibit an increase.
In general there are no evidence that indicates BTP Italia and Futura emission impact government bonds networks mean degree. As shown in the pictures before, mean degree in some cases appear to decrease, while in other situations it increases before, during and after BTP emission dates.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

Through this project we extract, study and analyze post-trade data from Monte Titoli, applying social network analysis. We extract the data from T2S system and use pyspark to aggregate data by day. After extraction, the datasets are stored in Amazon Simple Storage Service (S3), precisely in a bucket, to be later loaded in Amazon Sagemaker. Once data are loaded in Sagemaker machine, it have been pre-processed before the networks construction. The data are aggregated monthly and daily, specifically grouped by deliverer and receiver based on financial instruments such as ETF, government bonds, corporate bonds, funds, share and other. In addition, a refinement on instruction status is performed, filtering by failed and settled instructions. Social network graphs are constructed daily and monthly using NetworkX python library. The strategy is to simplify the instructions exchanged between companies by constructing graphs in which nodes represent companies and links represent instruction trades.

The links are weighted using the cash flow. Several analysis are performed for daily and monthly networks:

- Monthly Graph (cumulative and non-cumulative): inspectional analysis on the overall structure of the networks, centrality analysis, node importance and scale-free pattern detection and networks resiliency.

- Daily Graph: analysis of disruptive events.Two case are considered: the impact of Covid19 and BTP emissions on networks topology.

Preliminary inspection of the networks shows that nodes range from 30 to 346 depending on the financial instrument and settlement status. Government bonds networks with only-settle instructions are found to have more links and nodes with respect to networks with only-failed. In addition, a deeper analysis of the assortativity, average clustering coefficient, transitivity and density. The networks exhibit negative assortativity, large average clustering, moderate transitivity and moderate density. Furthermore, a connected component analysis was performed. The totality of the networks are found to be weakly connected, with the exception of the ETF only-settled network that results strongly connected

The detection of central nodes is performed according to Page Rank. We iterate the totality of the monthly networks and obtain a frequency score of appearing in the top-10 most central nodes. Cumulative monthly networks and noncumulative monthly networks report different results. Considering the non-cumulative monthly networks, South Construction, Net Venture, Cassa Compensazione e Garanzia, Contract Bell and Application Data Resource are the most frequent nodes, while considering cumulative networks, South Construction, Net Venture, Research System Design, Contract Bell and Cassa Compensazione e Garanzia are the most frequent. Several companies result frequent in both cumulative and non-cumulative networks. Furthermore, an analysis on monthly top-10 most central nodes is performed. Cumulative central nodes shows a stable pattern in nodes ranking, however non-cumulative central nodes do not exhibit any recognizable behaviors.

An in-depth the scale-free analysis detection in the networks is performed. Monthly cumulative networks do not exhibit Strongest, Strong and Weak scale-free behavioral, while 387 networks result Weakest Scale-free, 79 Super-Weak Scale-free and 69 Networks not Scale-free. Monthly non-cumulative networks do not exhibit Strongest, Strong and Weak scale-free behaviors, while 294 networks are found to be Weakest Scale-free, 89 Super-Weak Scale-free and 129 Networks not Scale-free. In general, cumulative networks show a predominance of scale-free behavior with respect to non-cumulative. Moreover, as monthly cumulative networks begin to expand, the number of scale-free networks also increases. This means that Monte Titoli's networks can be represented and modelled with a acceptable level of confidence using a power law distribution.
A study of network resilience is performed. Nodes are iteratively deleted and a frequency score is computed based on resulting damages in the network. After node is deleted, if the network shows isolated components, then the network is damage and the node is considered vulnerable. In particular, for cumulative networks North Source, Monte Titoli, Building Application Innovation, Cassa Compensazione e Garanzia and Internet East Hardware result in the most fragile nodes, while for non-cumulative networks North Source, Internet East Hardware, Cassa Compensazione e Garanzia, Monte Titoli and Start Power are the most fragile. Furthermore, a study on the change in topology after deletion is done. Eliminating an important node has a greater impact on the network topology than randomly eliminating a node according to size and average shortest path length. In general, the networks considered are more vulnerable to attacks localized to central nodes, while they are more robust to random failures. Monte Titoli's system appear to be resistant to elimination since the most common type of elimination would be random. The reason is that central nodes represent medium-large companies which are commonly stable and robust to bankrupt or exit from the system.
For daily graphs, we study the impact of disruptive events on the topology of the networks. A Covid19 case study is performed: the focus is detecting changes in topology caused by Covid19 pandemic. Mean de-

gree, average clustering and slightly assortativity have changed since the start of the pandemic, in particular a negative trend can be seen. For transitivity no particular pattern is found. A case study on BTP is performed: the aim is to detect changes in the topology during and after the emission dates of a specific government bond called BTP. No particular conclusions can be drawn in this case, since general there are no evidence that indicates BTP Italia and Futura emission impacted government bonds networks.

4.2 Future Work

For future work, there are several possibilities. Similar experiments such as node centrality, scale-free behavior, and networks resiliency can be also performed on daily networks. In this way, it could be possible to find if there are any change in the results considering daily networks instead of monthly. Especially with regard to scale-free detection, it might be useful to study how the scale-free degree changes on a daily basis. In addiction, nodes elimination in daily networks may exhibit interesting results on specific companies daily. Other types of node elimination may be applied, such as those presented in the study *"Comparison of Failures and Attacks on Random and Scale-Free Networks"* in Section 2.4. Furthermore, applying the same procedure on data from other contexts may be interesting to understand if similar behaviors of post-trade data can be detected. On February 2022 settlement discipline went live. It introduced penalty on instructions with long time failed status. It could be interesting to apply the same impact analysis of coronavirus and BTP but in this case considering daily networks with only-failed instructions before and after February 1^{st} 2022.

Bibliography

[1] Bank of Italy. Securities settlement systems.

[2] European Central Bank. Target2 calendar.

[3] Bank of Italy. Central counterparty. *Market and Payment System Oversight*.

[4] Mingxin Zhang. *Social Network Analysis: History, Concepts, and Research*, pages 3–21. Springer US, Boston, MA, 2010.

[5] D.N. Levine and G. Simmel. *Georg Simmel on Individuality and Social Forms*. Heritage of Sociology Series. University of Chicago Press, 1972.

[6] Barry Wellman. *The Development of Social Network Analysis: A Study in the Sociology of Science*, volume vol. 37. May 2008.

[7] Angela et al. Tringali. Life history stage explains behavior in a social network before and during the early breeding season in a cooperatively breeding bird. *PeerJ*, 2020.

[8] Rakhyun E Kim. Is Global Governance Fragmented, Polycentric, or Complex? The State of the Art of the Network Approach. *International Studies Review*, 22(4):903–931, 09 2019.

[9] Using social network analysis to understand missouri's system of public health emergency planners. *Public health reports*, vol. 122,4:488–98, 2007.

[10] *Dynamics On and Of Complex Networks III Machine Learning and Statistical Physics Approaches.* Springer proceedings in complexity. Springer International Publishing., 2019.

[11] Andrea Fronzetti Colladon and Elisa Remondi. Using social network analysis to prevent money laundering. *Expert Systems with Applications*, 67:49–58, jan 2017.

[12] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.

[13] Euronext. Euronext clearing.

[14] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nat Commun*, 2019.

[15] Stephanie Rendòn de la Torr, Jaan Kald, Robert Kit, and Juri Engelbrecht. On the topologic structure of economic complex networks: empirical evidence from large scale payment network of estonia. *Chaos Solitons*, 2016.

[16] Kimmo Soramak, Morten L. Bech, Jeffrey Arnold, Robert J. Glass, and Walter E. Beyeler. The topology of interbank payment flows. *Federal Reserve Bank of New York Staff Report*, 2006.

[17] Jean-Loup Guillaume, Matthieu Latapy, and Clémence Magnien. Comparison of Failures and Attacks on Random and Scale-Free Networks. pages 186–196. Springer-Verlag GmbH, 2004.

[18] Bank of Italy. Target2-securities: settlement of securities transactions.

[19] Apache. Pyspark. Apache. Available at `https://spark.apache.org/docs/latest/api/python/`.

[20] Sajee Mathew. Overview of amazon web services. 2014.

[21] Pieter Swar Aric Hagberg, Dan Schult. Networkx. Github Repository. Available at `https://networkx.org/`.

[22] Government bond. *Investopedia.*

[23] Corporate bond. *Investopedia.*

[24] Shares. *Investopedia.*

[25] Mutual fund. *Investopedia.*

[26] Exchange-traded fund (etf) guide. *Investopedia.*

[27] David N Fisher, Matthew J Silk, and Daniel W Franks. The perceived assortativity of social networks: Methodological problems and solutions, 2017.

[28] Michele Catanzaro, Mariá n Boguñá, and Romualdo Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Physical Review E*, 71(2), feb 2005.

[29] Jean-Paul Rodrigue, Claude Comtois, and Brian Slack. *Appendix A: Methods in Transport Geography.* 5th edition, May 2020.

[30] Pin Luarn and Yu-Ping Chiu. Influence of network density on information diffusion on social network sites: The mediating effects of transmitter activity. *Information Development*, 32(3):389–397, 2016.

[31] Andrew Disney. Social network analysis 101: centrality measures explained. 2020.

[32] Junlong Zhang and Yu Luo. Degree centrality, betweenness centrality, and closeness centrality in social network. *Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*, Vol. 132:p. 300–303, 2017.

[33] Howard T. Welser, Nina Cesare, Derek Hansen, and Md. Mahbub Or Rahman Bhuyan. Chapter 14 - wiki networks: Connections of culture and collaboration. In Derek L. Hansen, Ben Shneiderman, Marc A. Smith, and Itai Himelboim, editors, *Analyzing Social Media*

Networks with NodeXL (Second Edition), pages 205–224. Morgan Kaufmann, second edition edition, 2020.

[34] Derek L. Hansen, Ben Shneiderman, Marc A. Smith, and Itai Himelboim. Chapter 3 - social network analysis: Measuring, mapping, and modeling collections of connections. In Derek L. Hansen, Ben Shneiderman, Marc A. Smith, and Itai Himelboim, editors, *Analyzing Social Media Networks with NodeXL (Second Edition)*, pages 31–51. Morgan Kaufmann, second edition edition, 2020.

[35] Daron Acemoglu, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80:1977–2016, 2012.

[36] Anna Broido. Sfanalysis. Github Repository. Available at `https://github.com/adbroido/SFAnalysis`.

[37] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, jul 2000.

[38] Peterson Ozili and Thankom Arun. Spillover of covid-19: Impact on the global economy. Mpra paper, University Library of Munich, Germany, 2020.

[39] Daniele Palumbo Lora Jones and David Brown. Coronavirus: How the pandemic has changed the world economy. *BBC News*, January 2021.

[40] FTA Online News. I btp: cosa sono, quanto rendono e come acquistarli [btps: what they are, how much they yield and how to buy them]. *FTA Online News*, July 2021.

[41] Italian Ministry of Economy and Finance. Treasury bonds step-up (btp futura). *Italian Ministry of Economy and Finance.*