

# Networks of shares: SP500 from 1962 to 2022

Manuel Luci

m.luci@studenti.unipi.it

Student ID: 581770

Ayushy Vijaykuma

a.vijaykumar@studenti.unipi.it

Student ID:

Alessandro Seccarelli

a.seccarelli@studenti.unipi.it

Student ID: 583814

## ABSTRACT

In this project, we aim to study the stock market with the tools of graph theory. In particular, we will study the static and dynamic properties of the graph obtained by considering as nodes the individual stocks included in the SP500 index, and as adjacency relation being strongly correlated. After a quick statistical study of the general topological characteristics of the graph obtained by considering the entire temporal span (1962-2022), we will devote attention to community discovery, both in the dynamic and static case, attempting to make identifications with the ground truth offered to us by the sectoral affiliation of the various firms. The results on this issue tend to be negative, as it is not possible to draw conclusions that support the initial hypothesis of coincidence between communities and sectoral grouping. Finally, we will investigate whether and how the major events disrupting the world's stock exchanges can be traced in our data: above all, the two most recent financial crushes (subprime mortgage crisis and the global pandemic).

<sup>1</sup>

## KEYWORDS

Social Network Analysis, Stock market network, Financial crushes, Speculative bubble, Communities

## ACM Reference Format:

Manuel Luci, Ayushy Vijaykuma, and Alessandro Seccarelli. 2019. Networks of shares: SP500 from 1962 to 2022. In *Social Network Analysis '22*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

## <sup>1</sup>Project Repositories

Data Collection: [https://github.com/sna-unipi/sna-project-2022\\_luci\\_seccarelli\\_ayushi/tree/main/data\\_collection](https://github.com/sna-unipi/sna-project-2022_luci_seccarelli_ayushi/tree/main/data_collection)

Analytical Tasks: <https://github.com/sna-unipi/analytical-tasks>

Report: [https://github.com/sna-unipi/sna-project-2022\\_luci\\_seccarelli\\_ayushi/tree/main/report](https://github.com/sna-unipi/sna-project-2022_luci_seccarelli_ayushi/tree/main/report)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. SNA '22, 2021/22, University of Pisa, Italy

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$0.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

*Analysis '22*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

## 1 INTRODUCTION

Social Network analysis is a useful tool to analyse the stock market: the graph has firm as nodes and "strength" of correlation as edge. In the following project we have considered only strong correlation (Pearson correlation  $> 0.8$ ) and negative one (Pearson correlation  $< -0.5$ ).

## 2 DATA COLLECTION

As source we have considered Yahoo Finance and through the package YFINANCE we get annual historic data for each firm of S&P 500 since 1962, we saved it in a dictionary where each key is a stock (**dataset\_building.py**). We build a yearly correlation matrix (evaluate pearson correlation of adjusted closed price) and saved them in a dictionary, each key is the considered year, we cleaned them (deleting NaN values and 0 values), (**correlation\_building**, **cleaning\_correlation**, **correlation\_building\_aggregate**).

As last step we built graph using this data (node: stocks, edge: correlation, label : sector): **list\_pos\_graph\_label.npy** is a dictionary of strongly correlated stocks, **list\_neg\_graph\_label.npy** is a dictionary of inversely correlated stocks, each key is a year; **pos\_graph\_tot.pkl** and **neg\_graph\_tot.pkl** are graph strongly and negatively correlated obtained considering the entire timeline (**graph\_building**, **graph\_splitting**, **label\_graph**).

## 3 NETWORK ANALYSIS

### Static Network

First of all we analyzed statistical behaviour of static network (we call positive = strongly correlated network, negative = weakly correlated network); both positive and negative have 494 nodes, but positive has 70567 edges, negative 1026, the average degree of positive is 285.70 while negative is 4.15. Studying the distribution of degree gives useful information about graph's topology: plotting both distribution in loglog scale shows a power law behaviour for the negative one: Fitting a powerlaw distribution  $y = \alpha + \beta x$  ( $y$  = number of nodes,  $x$  =  $\log(\text{degree})$ ), we obtain:  $\alpha = 5.43 \pm 0.30$ ,  $\beta = -1.65 \pm 0.11$ ,  $R^2 = 0.87$ ,  $\chi^2 = 0.99$ , this result is in agreement with general financial theory.

Diameter of biggest component of positive is 5, for negative 6 and avarage shortest path are respectively 1.47 and 2.74.

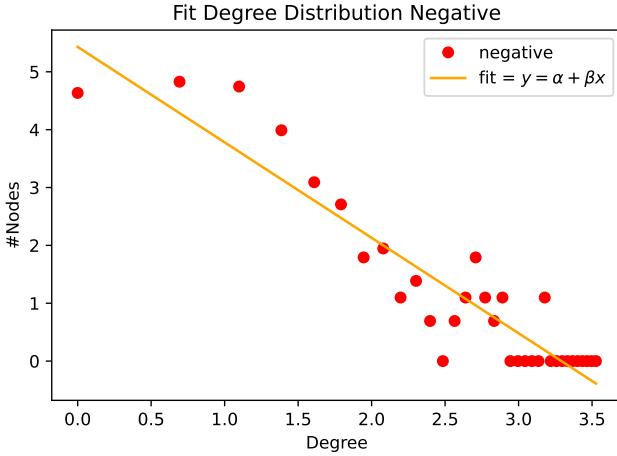


Figure 1: Degree Distribution negative graph fit

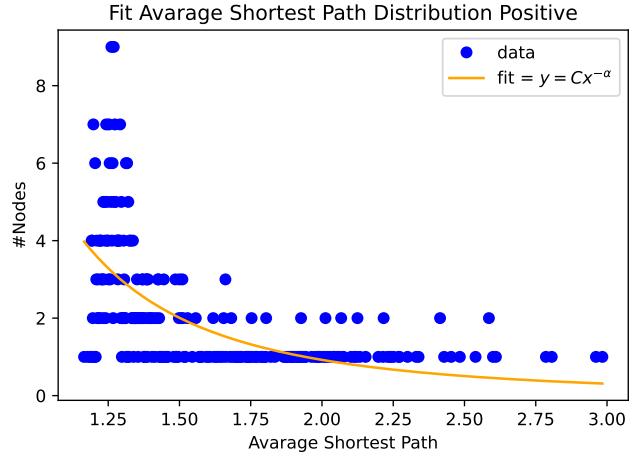


Figure 3: Average Shortest Path Distribution for Positive

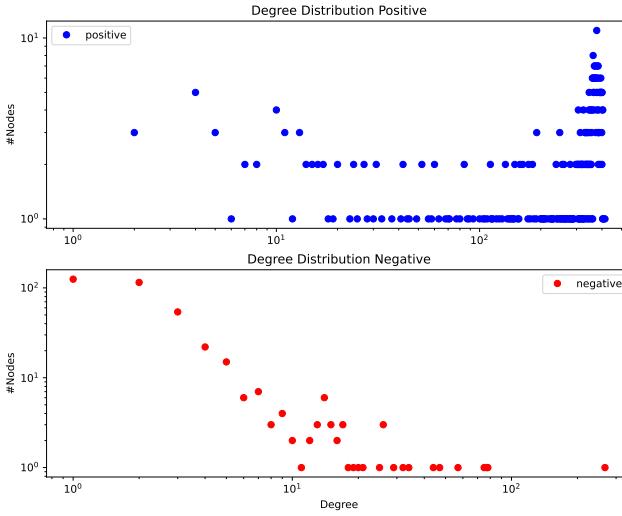


Figure 2: Degree Distribution positive and negative graph

Plotting average shortest path distribution for positive, shows powerlaw behaviour, fitting this distribution with  $y = Cx^{-\alpha}$  we obtained:  $C = 6.01 \pm 0.55$ ,  $\alpha = 2.71 \pm 0.29$ ,  $R^2 = 0.36$ ,  $\chi^2 = 172.81$ . Density of positive and negative is 0.58 and 0.01, while global clustering coefficient is 0.834 and 0.001.

**Comparision with ER.** A ER graph having same edges and nodes of our positive and negative graph need  $p = 0.58$  for positive and 0.01 for negative, positive is in connected regime, negative in supercritical; plotting positive and negative degree distribution we can notice as ER graph is not a good model for our graphs.

**Comparision with BA.** A BA graph having same nodes and edges of our positive and negative graphs need  $m$ (number

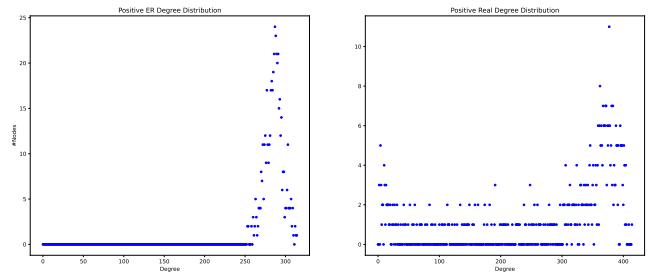


Figure 4: Positive ER Degree Distribution

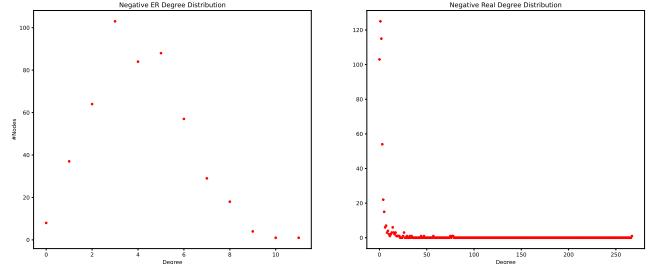


Figure 5: Negative ER Degree Distribution

of edges to attach) = 143 for positive, 2 for negative. Fitting degree distribution with a powerlaw we obtained for positive:  $\alpha_{BA} = 5.29 \pm 0.41$ ,  $\alpha_{Real} = 15.38 \pm 5.06$ , for negative:  $\alpha_{BA} = 3.67 \pm 0.20$ ,  $\alpha_{Real} = 7.05 \pm 0.31$ , BA graph with this parameters is not an appropriate model for our graphs as well.

**Closeness centrality.** By evaluating closeness centrality we obtained the following results: positive graph closeness values are higher than negative one, considering the 20 largest values of closeness for positive and negative, we noticed that

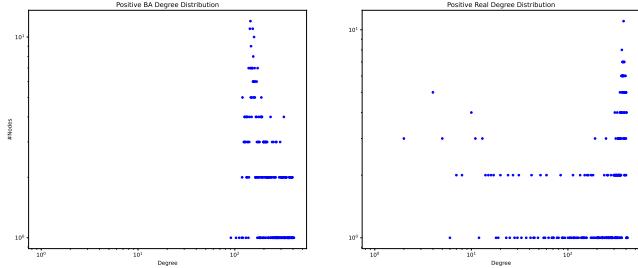


Figure 6: Positive BA Degree Distribution

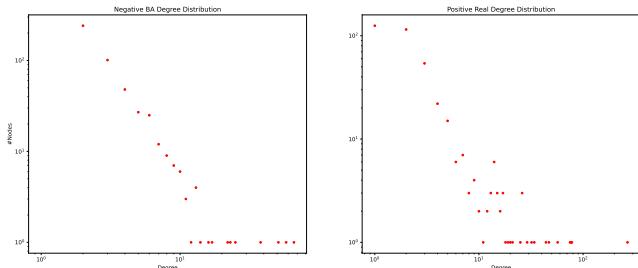


Figure 7: Negative BA Degree Distribution

30% of stocks belongs to Industrial sector for positive and 30% belongs to Information Technology for negative.

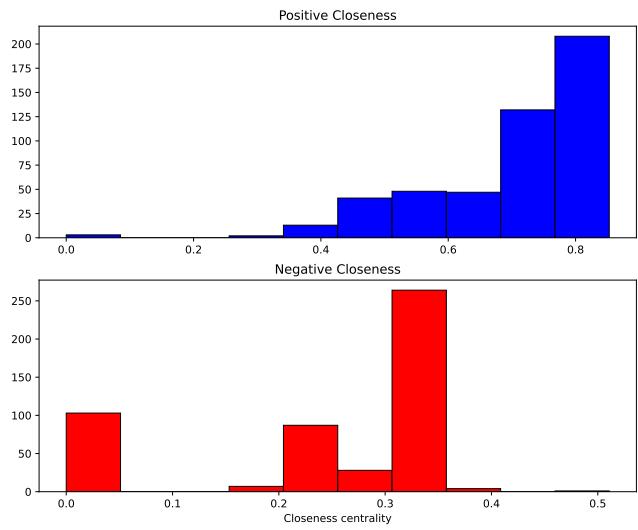


Figure 8: Distribution Closeness

*Betweenness centrality.* Evaluating betweenness centrality we obtained the follow results: most values for positive are nearly 0.0025 and 0.05 for negative, these small values are due the absence of meaningful bridges in the graph; considering the biggest 20 values we noticed as 30% of stocks belongs to Financials sector for positive and 25% belongs to Information

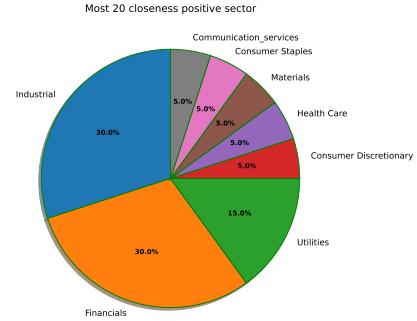
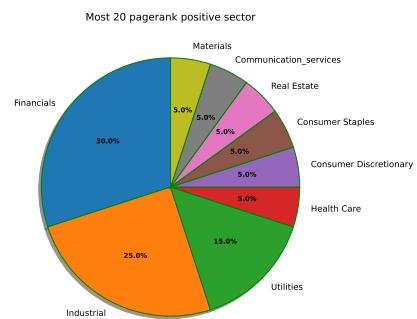


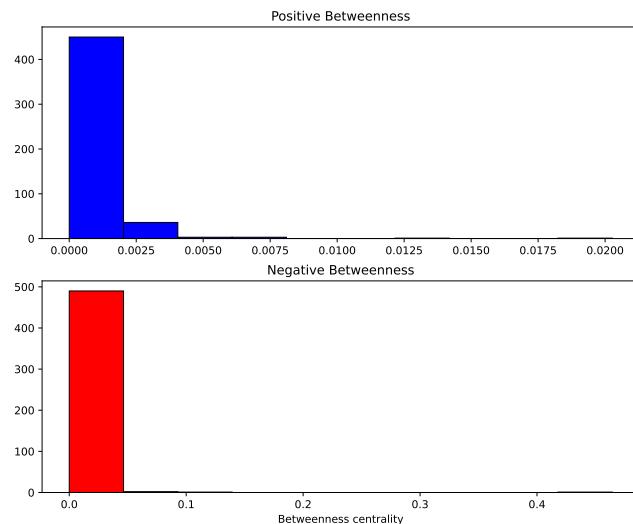
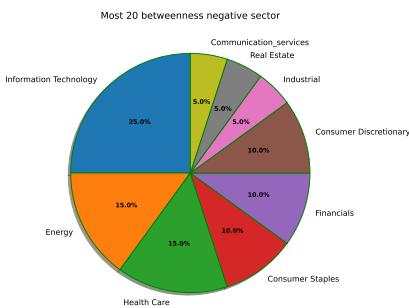
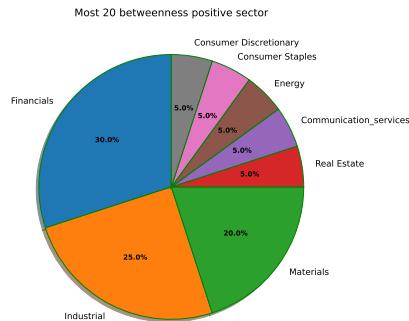
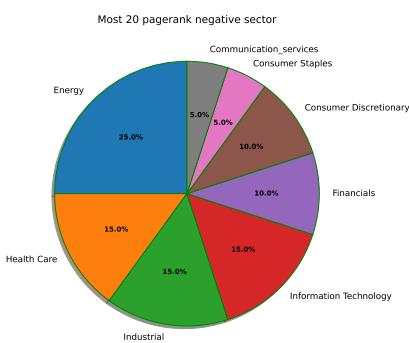
Figure 9: Pie Chart Closeness



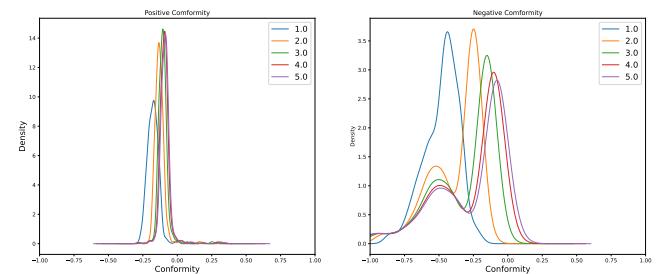
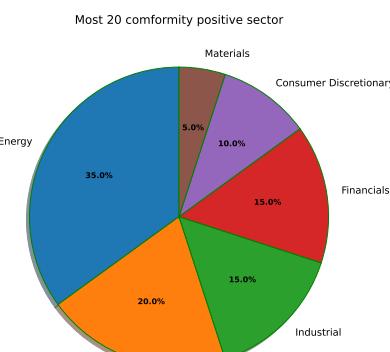
Figure 10: Wordcloud Closeness

Technology for negative. Evaluating pagerank, considering the biggest 20 values we noticed as 30% of stocks belongs to Financials sector for positive and 25% belongs to Energy for negative. This results show the great centrality of the financial sector in stock market.



**Figure 11: Distribution Betweenness****Figure 12: Pie Chart Betweenness****Figure 13: Pie Chart PageRank****Figure 14: Wordcloud PageRank**

*Assortativity measures.* Considering sectors, positive's assortativity is 0.01, while negative's assortativity is -0.08, both graphs do not show homophily behaviour. Using conformity with a decay factor of 3, 35% of the least 20 disassortative firms belong to Energy

**Figure 15: Conformity at different decay values**

#### 4 TASK 1: DYNAMIC NETWORK ANALYSIS

We built a dynamic graph considering 60 snapshots (from 1962 to 2021)

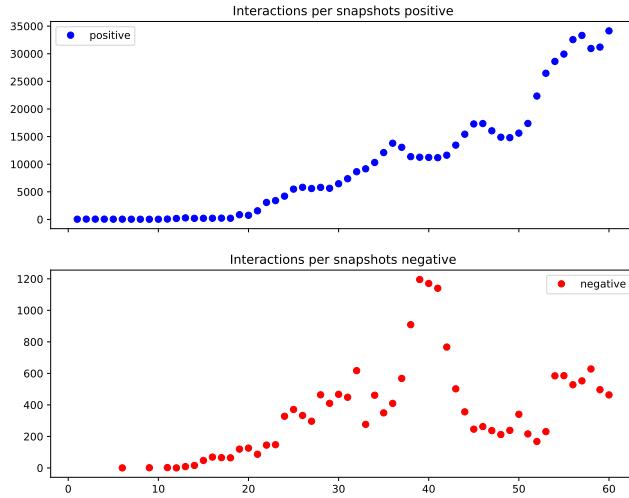


Figure 16: Interaction per snapshots

Considering the interaction per snapshots values, positive graph tend to increase, while negative have an unstable behaviour.

Studying the distribution of degree over snapshot, we noticed that the degree distribution tends to stabilize, we fitted the degree distribution of the 60th snapshot of the negative graph with a power law ( $Y = A + BX$ , data are in logs) obtaining:  $A = 3.98 \pm 0.34$ ,  $B = -1.03 \pm 0.11$ ,  $\chi^2 = 0.99$ ,  $R^2 = 0.74$ . Coverage for the positive graph is 0.55, while for the negative one is 0.50. By studying the contribution of each node in both graph, the top 3 nodes associated to the highest values in contribution belong to Energy sector for both graph. The density of the positive graph tends to increase, albeit with significant fluctuations snapshot by snapshot, while that of the negative converges to 0.

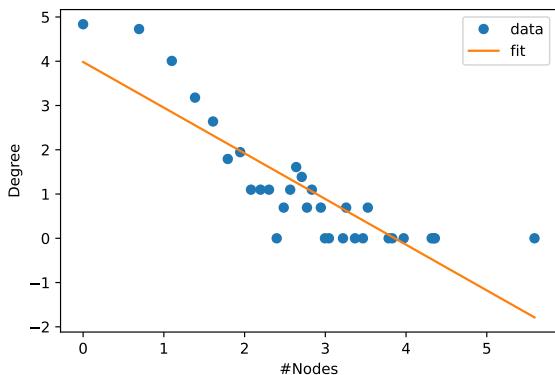


Figure 17: Fit 60th snapshot negative degree distribution

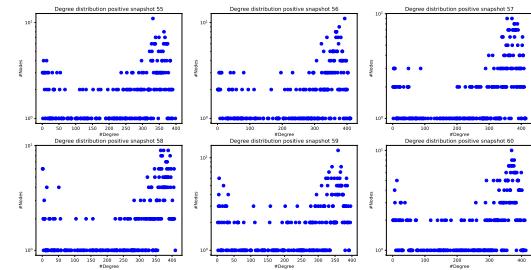


Figure 18: Snapshot positive degree distribution

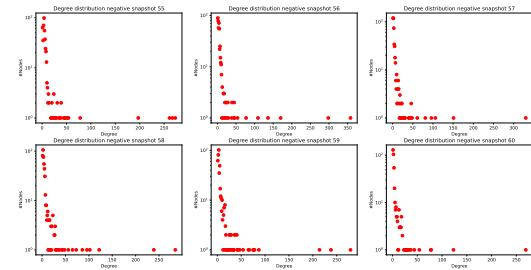


Figure 19: Snapshot negative degree distribution



Figure 20: Wordcloud Contribution

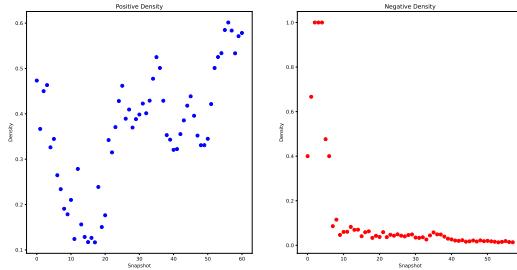


Figure 21: Density over snapshot

## 5 TASK 2: STATIC COMMUNITY DISCOVERY

In this section we focus on the detection of static community discovery. To do so, we will study the overall graph, `pos_graph_tot.gpickle`. We have applied four different static community discovery algorithms (Leiden, Girvan-Newman, Newman's leading eigenvector, Label propagation) to understand how various types of methods interact with our financial network. We begin by analysing the two cases in which the result is negative, no relevant partitions is returned : Girvan-Newmann and Label propagation. In the first case, it appears that, although the dendrogram is interrupted at various levels (in particular up to 30), there is only one large central cluster with many communities of very small size ( $< 3$  nodes). The explanation is to be found in the high density of the network and its relative compactness due to the strong global trend component that characterises stock markets: subtracting even a significant fraction of the links leads to separating nothing but few shares. It is an example of how the two heuristic ideas of community partitioning (density vs bridges) do not always lead to similar results, particularly for very dense graphs in which the identification of bridges has little meaning.

The label propagation algorithm, on the other hand, leads to an opposite result, albeit for the same reasons: since this is based on a bottom-up approach instead of a top-down one, the high density and compactness of the network does not allow for the aggregation of communities: each node remains separate.

The two remaining algorithms, Leiden and Newman's leading eigenvector, lead to results that are a little more interesting: both identify two large clusters, which are, moreover, mutually consistent as we shall see later on, which is partly to be expected given that both are linked to the use of modularity as a value to be maximised. The coincidence between the two is however of interest given that they present two quite different methods: Leiden's algorithm has a bottom-up approach with a standard greedy procedure, while Newman's leading eigenvector algorithm is top-down, algebraic one.

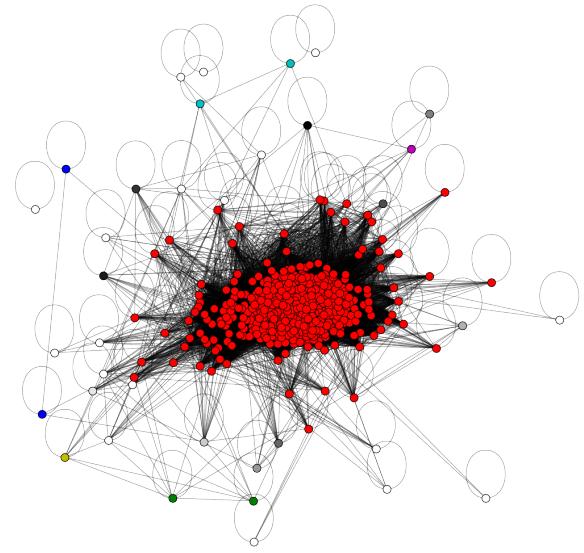


Figure 22: CD with GN algorithm at level 30, graphical representation

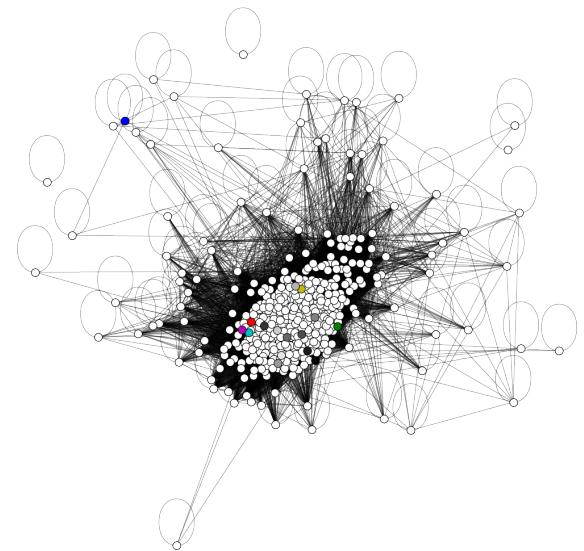
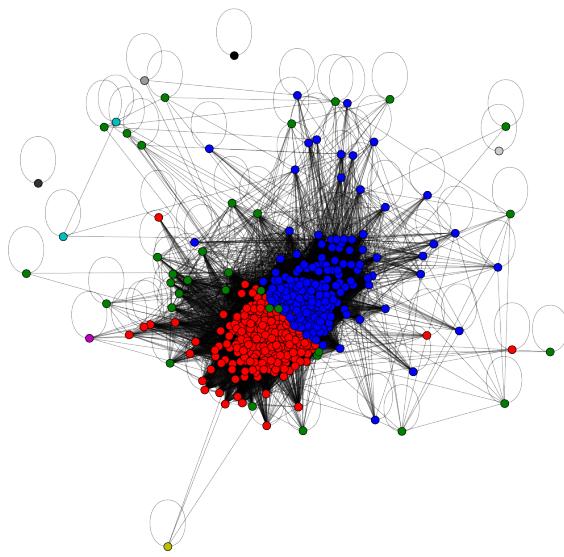
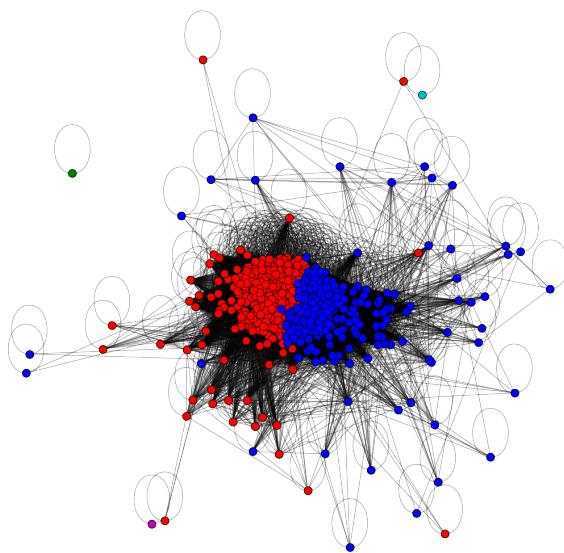


Figure 23: CD with LP algorithm, graphical representation. White nodes are not a

With these two methods, based on the maximisation of an objective function, one does not run into the difficulties seen above due to the compactness of the network.



**Figure 24: CD with Leiden algorithm, graphical representation.**



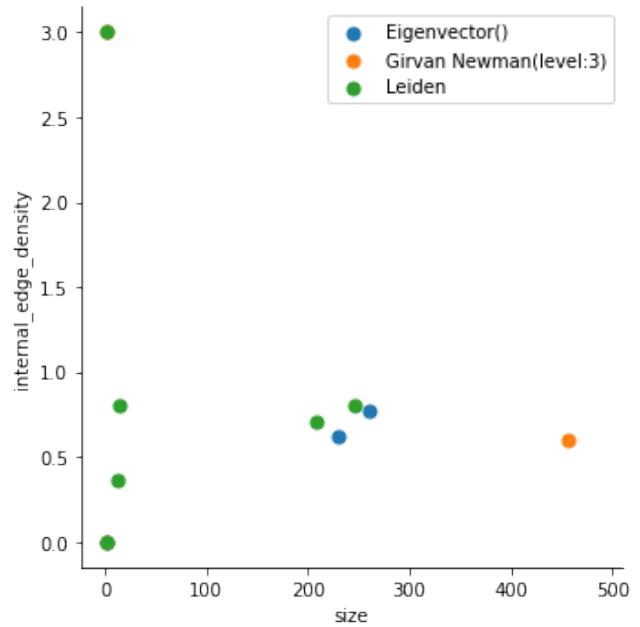
**Figure 25: CD with Newman's leading eigenvector algorithm, graphical representation.**

We employ both internal and external measures as evaluation tools. Of the former type, we are interested in modularity and conductance. Of the second, we will compare both the various partitions with each other and with the ground truth

offered to us by the sectoral classification of listed companies. The following table summarises the results of the two statistics:

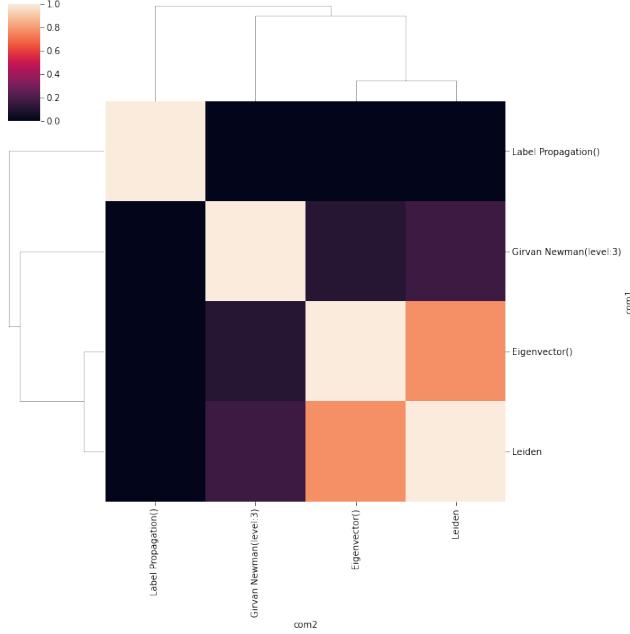
Algorithm	Modularity	Conductance
Leiden	0.0969	0.4445
Leading eigenvector	0.0962	0.4071
Girvan-Newman	0.0016	0.6586
Label propagation	0.0052	0.9677

Even the two algorithms based on modularity do not perform satisfactorily with this statistic, among other things producing a result of similar quality. This is probably again attributable to the great compactness of the graph. The other proposed internal evaluation measure also identifies Louvain and Leading eigenvector as higher quality partitions. Finally, from the following graph (26) we can get an idea of the internal density for each of the various communities identified through the various algorithms as a function of their respective size (number of nodes).



**Figure 26: Internal density vs size of community. Label propagation not represented**

Again, In order to compare the various partitions obtained, we employ adjusted mutual information, which is actually more useful in quantifying how correct is the idea that the Leiden algorithm and Newman's leading eigenvector algorithm actually provide similar responses. Indeed, only for this pair of partitions does the index in question take on a significant value, and in particular it is worth around 0.73. The similarity matrix summarises these comparisons.



**Figure 27: Similarity matrix with adjusted mutual information. Clearly only the communities detected with Leiden and Newman's leading eigenvector algorithm are coherent**

The same index was also used for the ground truth comparison: using the GICS (Global Industry Classification Standard, the industrial taxonomy employed by StandardPoor's, the corporation that builds the SP500 index) a partition of 11 communities was constructed based precisely on the sectoral affiliation of each company. Given the great disparity in number and relative size, intuition suggests that agreement will be rather poor. In fact, ignoring the two algorithms that produced no useful results, Leiden and Newman's leading eigenvector algorithms are also rated 0.0719 and 0.0591 respectively. The idea behind this comparison is that companies belonging to the same sector are subject to similar exogenous shocks, thus experiencing greater correlation in their respective share prices and thus being 'candidates' for clusters. The failure of this hypothesis may be due to several factors: for instance, it may be that the influence of sectoral shocks is limited when compared to shocks of another nature, such as the general business cycle or geographical shocks. Alternatively, it could be the choice of algorithms employed that are not appropriate. Or again, even if the sectoral component is significant in determining correlations between stock price movements, the chosen threshold of 0.7 is not sufficient to filter out larger scale components (again, the business cycle, even of distinct geographic areas). The high density of the network is a clue, leaning in this case towards the third, or possibly the first, hypothesis.

## 6 TASK 3: DYNAMIC COMMUNITY DISCOVERY

In this section, we will show, over a limited time interval (2010-2014), the evolution of communities in our financial network and the tools used to explore this aspect of the topic under investigation. We employed two different methods to reconstruct the history of communities: instant optimal and tiles. To these, we coupled the same algorithms used in the previous section to identify communities at a certain snapshot: as it previously turned out, the use of the label propagation algorithm has not proved effective for studying the behaviour of financial networks, so in this section we will not present the results for this community detection criterion, which in any case are inconclusive, as no community of relevant size has ever been circumscribed. Before presenting these results, we report the values of the temporal stability assessments of the identified communities, which are independent of the way the dynamic is reconstructed, obtained using the Normalized F1 score.

Alg.	year	10/11	11/12	12/13	13/14
Leiden		0.051	0.116	0.068	0.179
Lead. eigen.		0.200	0.059	0.055	0.062
GN		0.053	0.093	0.047	0.037

Let us now see the outcome of the instant optimal algorithm in conjunction with the GN algorithm: the graphs in the figure are polytrees that have as nodes the communities at a given time instant and as oriented arcs the identifications from one year to the next between the communities. With the algorithm, which has as input the jaccard index  $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$ , where  $A$  and  $B$  are sets), merges and splits of communities are determined. As Figure 33 makes clear, once the smaller communities have been eliminated, the evolution is rather limited: only one large community persists over the five years considered. Studying its size trend, we find that there is a tendency for its size to increase (from 439 to 464), due to the nodes being added to the financial network as they go along.

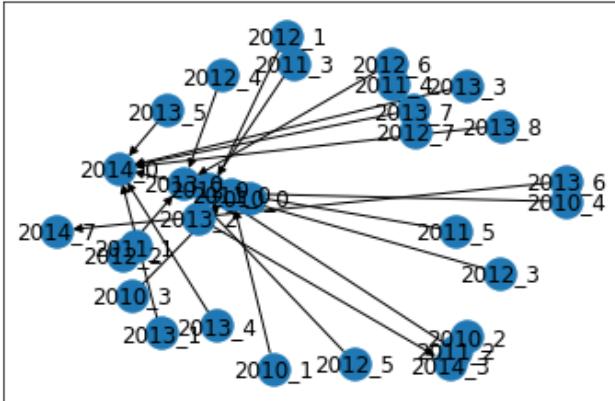


Figure 28: Polytree of dynamic community discovery with instant optimal algorithm, applied to communities discovered with GN algorithm.

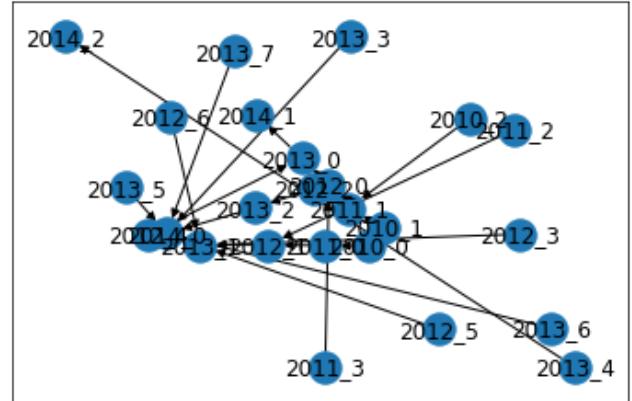


Figure 30: Polytree of dynamic community discovery with instant optimal algorithm, applied to communities discovered with Newman's leading eigenvector method.

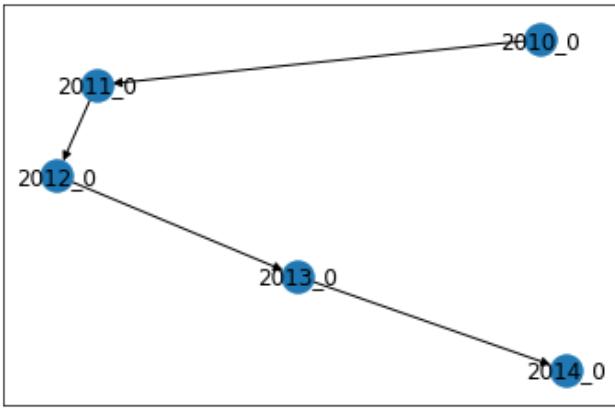


Figure 29: Polytree of dynamic community discovery with instant optimal algorithm, applied to communities discovered with GN algorithm. Small ( $< 5$  nodes) communities have been removed

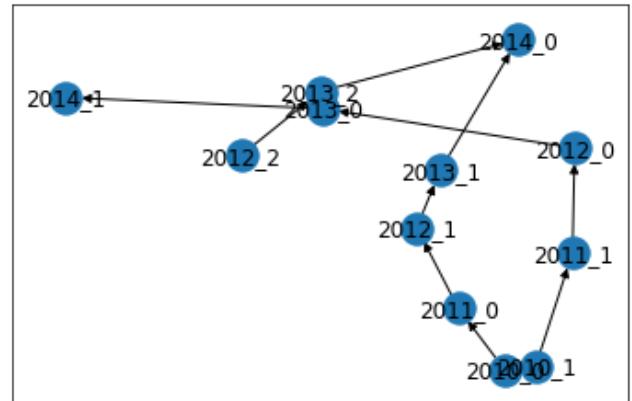
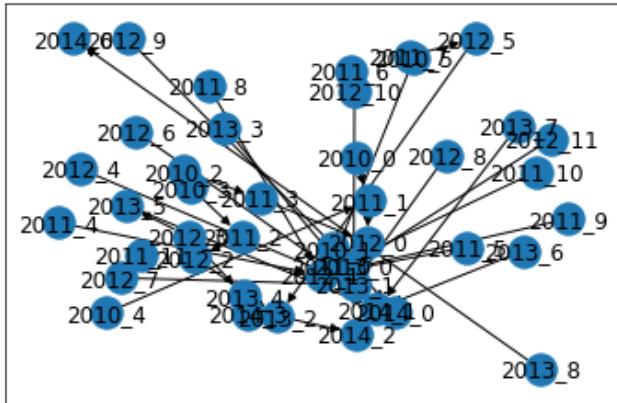


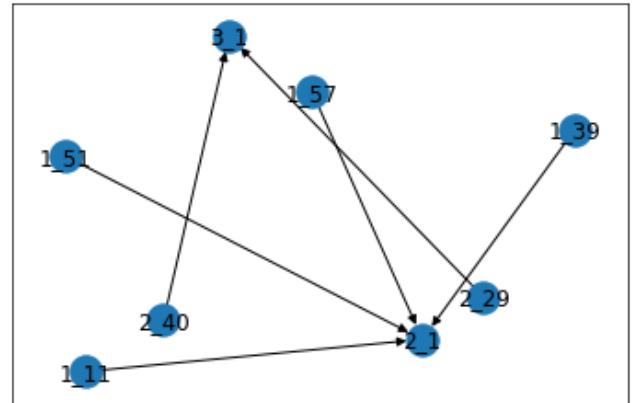
Figure 31: Polytree of dynamic community discovery with instant optimal algorithm, applied to communities discovered with Newman's leading eigenvector method. Small ( $< 5$  nodes) communities have been removed

Turning now to the results obtained with the GN algorithm, the story told by the 'cleaned-up' polytree of the smaller communities is richer: initially there are only two relevant communities, both stable with just over 200 nodes, in 2012 a third community with 100 nodes appears, while the larger of the two already existing communities loses about 80 members, only to reabsorb in 2014 'the newest arrival'.

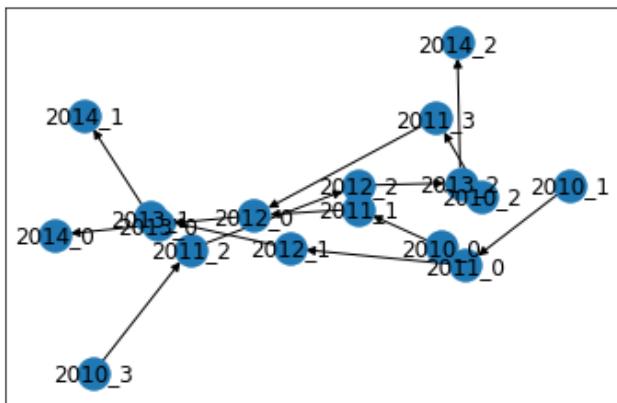
The last algorithm, Leiden, is undoubtedly the one that gives us the most complex case: in 2010 there are four different communities of significant size (171, 147, 89, 33 respectively). After some variation in size in 2011, in 2012 there is the merger of the largest with the smallest, while the others experience fluctuations in size.



**Figure 32:** Polytree of dynamic community discovery with instant optimal algorithm, applied to communities discovered with Leiden algorithm.



**Figure 34:** Polytree of dynamic community discovery with tiles method, applied to communities discovered with GN. Small (< 5 nodes) communities have been removed.

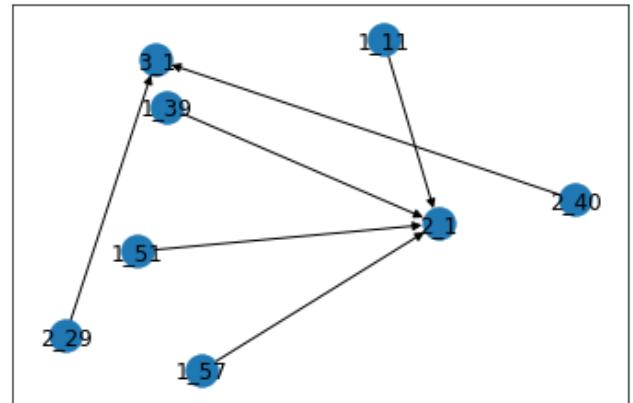


**Figure 33:** Polytree of dynamic community discovery with instant optimal algorithm, applied to communities discovered with Leiden algorithm. Small (< 5 nodes) communities have been removed.

The other method used to reconstruct the temporal evolution of communities places less emphasis on those identified in individual frames and prefers to build them incrementally as the network evolves. Because of this characteristic, the brevity of the analysed interval and the lack of stability and poor definition of the communities (e.g. we have just seen that the three previous algorithms return three different narratives), the results are unenlightening.

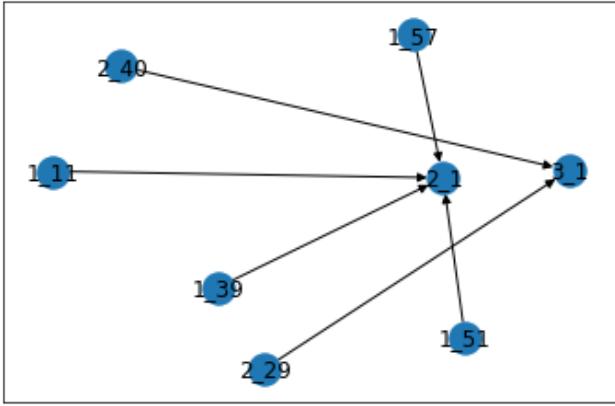
Going to interrogate the first polytree of this series of three, the one where communities are produced with GN algorithm: we observe that there are no communities from the last two snapshots. Beyond this, we observe two functions, the first of three entities and the second of two.

As in the previous one, in this second polytree obtained with Newman's leading eigenvector method, no communities were identified in the last two snapshots. Mergers are also recorded here.

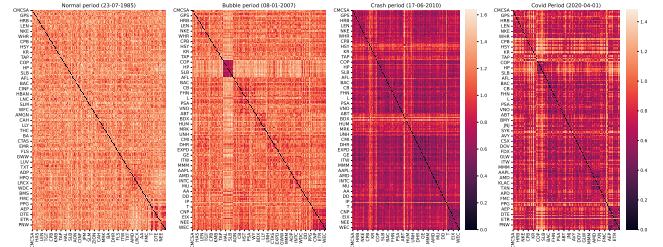


**Figure 35:** Polytree of dynamic community discovery with tiles method, applied to communities discovered with Newman's leading eigenvector method. Small (< 5 nodes) communities have been removed.

The last polytree in the series, where the communities were identified with the Leiden algorithm, has the same characteristics as the others.



**Figure 36: Polytree of dynamic community discovery with tiles method, applied to communities discovered with Leiden algorithm. Small ( $< 5$  nodes) communities have been removed.**



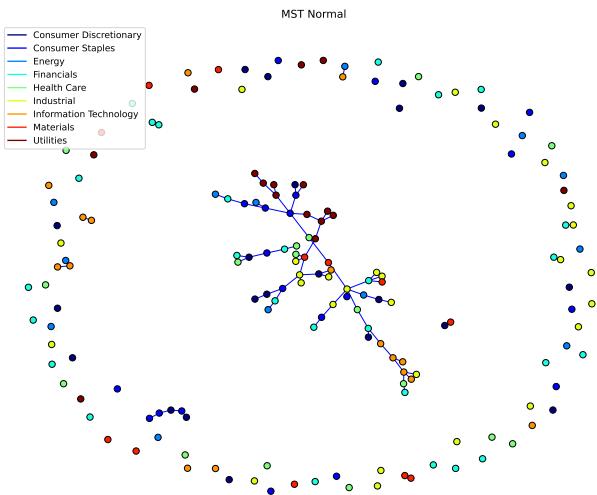
**Figure 37: Heatmap at various market periods**

## 7 TASK 4: STOCK MARKET CRUSH AND NETWORKS

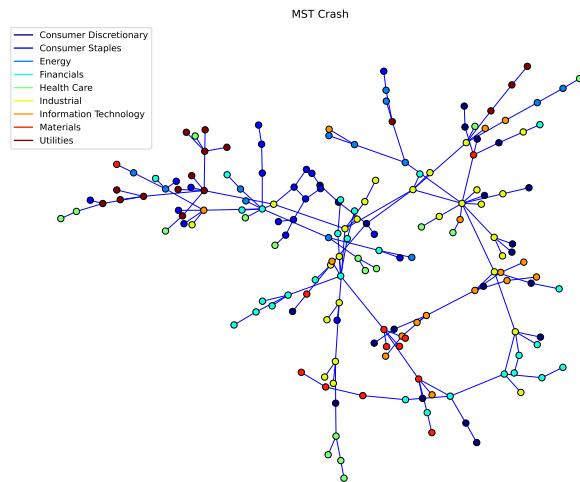
In this section, we try to answer the question that prompted the project: do graph theory tools have anything to say about times of major upheavals in the financial markets, such as mounting speculative bubbles and subsequent crashes, or even during intense shocks such as the collapse due to the Coronavirus pandemic? To do so, let us study how the main network-defined statistics behave during the various phases that stock markets experience: speculative boom and bust ('07-'10), March-April 2020 (pandemic shock) and a standard period (1985). In this section, correlations are calculated over a 40-day time interval, with rolling windows of 20 days. A further difference from the previous sections, here we study a weighted graph: each arc has assigned a value called distance calculated as follows:  $d = \sqrt{2(\rho - 1)}$ . Note that the higher the distance, the less strong the link between two stocks, as they are less correlated. The first tool used is the heat map: scrolling from left to right we encounter: in a normal period there are no sectors that stand out above the others. In the second grid, we can see that two sectors, nominally energy and IT sector, show a high level of clustering: this is an indicator of a possible bubble: a certain production sector is considered promising per se, without distinguishing between individual companies. When the moment of collapse arrives (third heat map), the previous hypothesis is verified, in particular the sectors whose share prices had been driven up by enthusiasm and not by rational evaluations are discovered. The collapse related to the pandemic shock resembles the crash period of the previous situation: after all, it was an event that negatively impacted most sectors.

Similar conclusions are also suggested to us by another useful tool for investigating the topology of the financial network and how it evolves: the Minimum spanning tree (MST) [1] offers a measure of how compact the stock market is overall. Again, in a period without turbulence, the financial network is sparse, unconnected and with few stocks experiencing interrelated price dynamics, so a few arcs suffice to make the minimum spanning tree. Something different happens progressively as one approaches a period of sharp declines in stock market indices: gradually more securities are aggregated from the central body of the minimum spanning tree and more edges are needed to realise it: stronger connections (i.e. higher correlation and lower distance) are gradually added, which make the various nodes even 'closer' and are thus selected by the greedy algorithm that produces the spanning tree. Again, we observe how the collapse due to the pandemic has similar characteristics to the speculative event of 2008, although it was clearly not preceded by a bubble. An advantageous feature of this tool is the clarity with which it provides an answer precisely because it only highlights the connections that are really necessary to illustrate how much the various prices of securities are moving together.

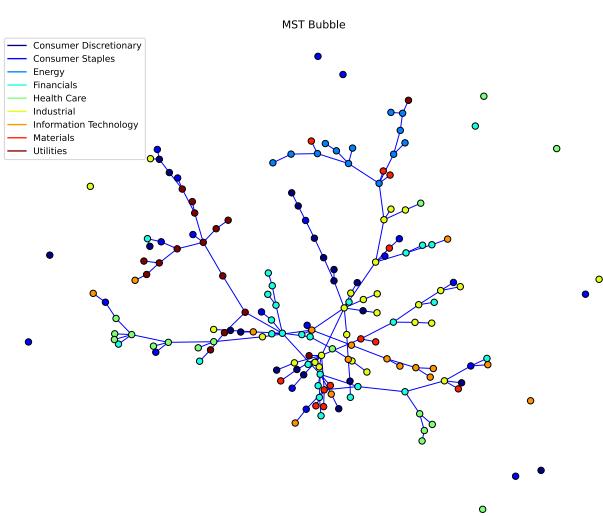
In order to have more information on the graph, it is possible to build the Planar Maximally Filtered Graph[2], it has more information about the connection than MST.



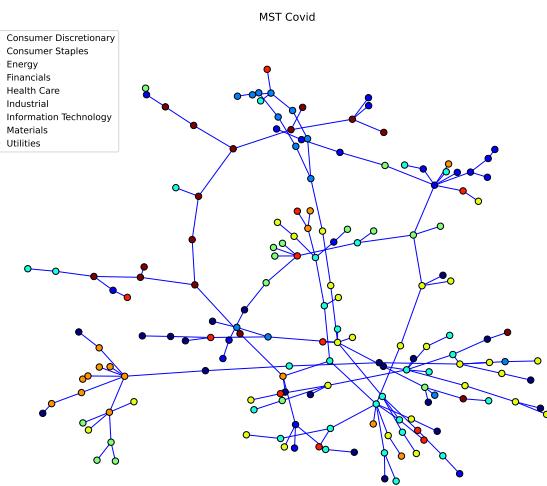
**Figure 38: Minimum spanning tree in normal market period (1985)**



**Figure 40: Minimum spanning tree in bust period (2010)**



**Figure 39: Minimum spanning tree in bubble period (2007)**



**Figure 41: Minimum spanning tree in covid crisis period (2020)**

The threshold network also suggests at a glance considerable qualitative differences between the normal state of the market and the more or less positive turmoil that invests it at other times: again, during the bubble's mounting, nodes begin to thicken as if preparing the bust's phase. Really a phase transition seems to be at work: we observe an evolution from an highly dispersed network to a(n almost) fully connected one.

A further analysis on threshold network have been done to discover some graph's topological properties useful to characterize the three period: Density and Global clustering are

first important values, an high value of density and Clustering Coefficient is symptom of crash period; the different shape of histogram of Local Clustering Coefficient gives us important information about Covid period: while during the Crash, almost all nodes have high values, for Covid the histogram have a left fat-tail shape, completely different from the previous ones.

Also betweenness centrality's during Covid period have a unique shape really different from previous one.

Assortative gives us another important parameter: in Crash and Covid period it is negative.

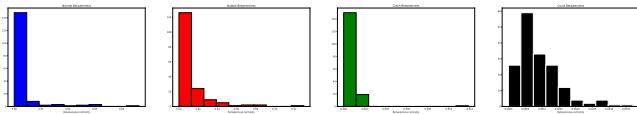


Figure 42: Betweenness centrality Histogram

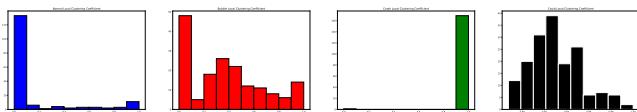


Figure 43: Local Clustering Histogram

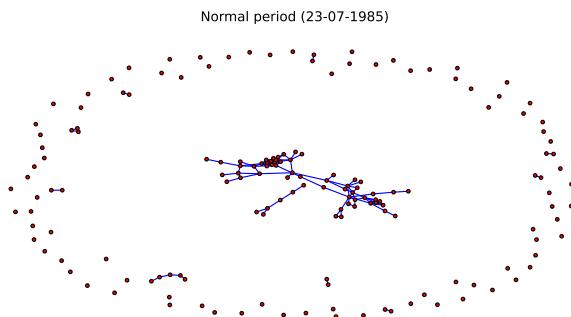


Figure 44: Threshold network with threshold  $d = 1$  in normal market time (1985)

Bubble period (08-01-2007)

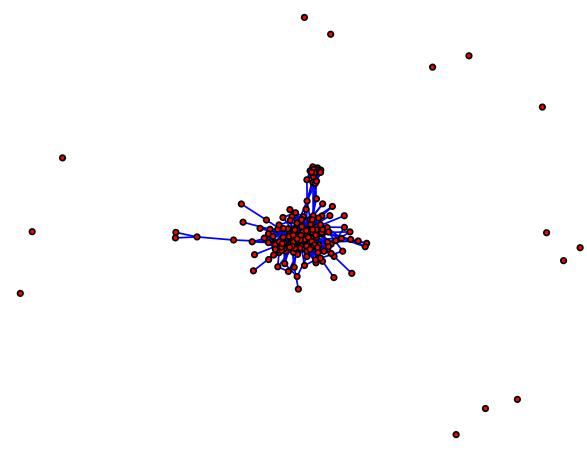


Figure 45: Threshold network with threshold  $d = 1$  in bubble period time (2007)

Crash period (17-06-2010)

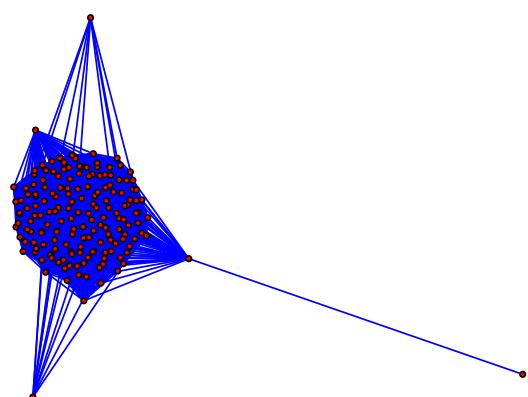
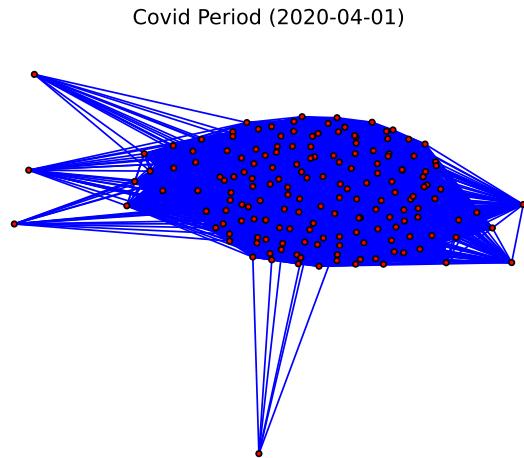
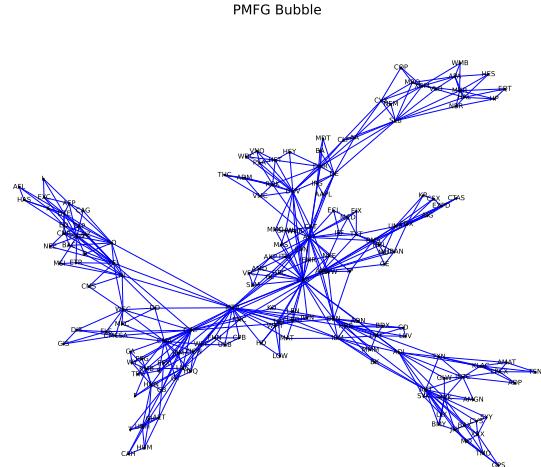


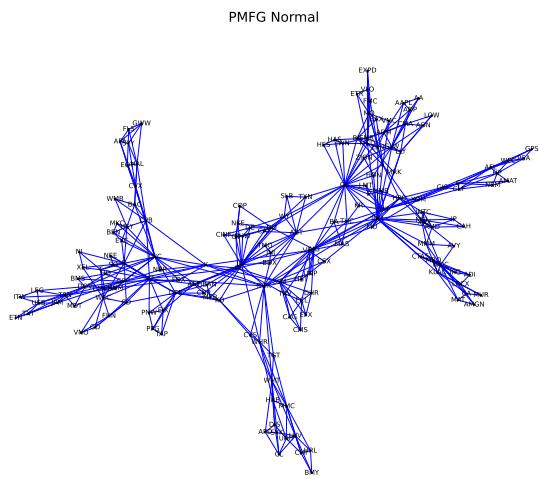
Figure 46: Threshold network with threshold  $d = 1$  in crash time (2010)



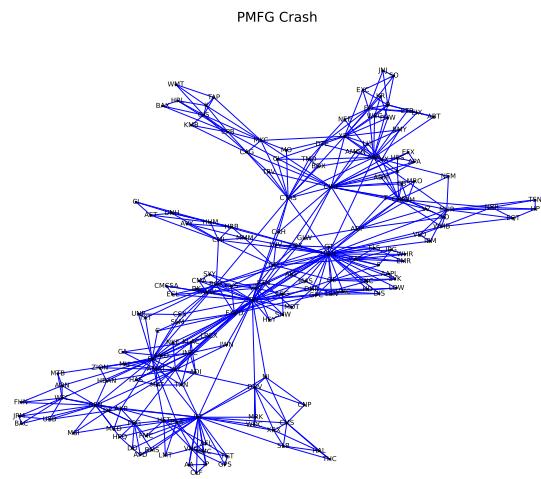
**Figure 47:** Threshold network with threshold  $d = 1$  in covid crisis (2020)



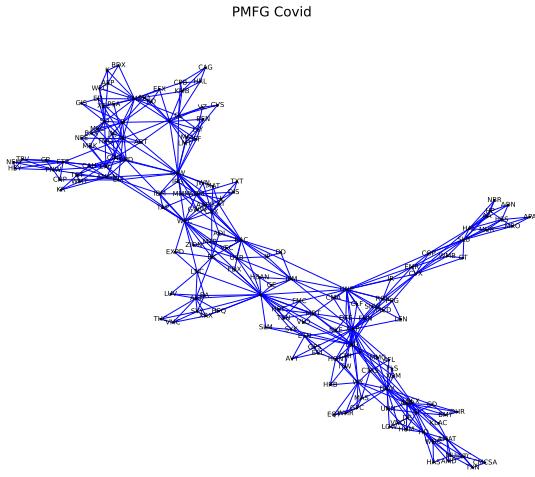
**Figure 49:** Planar maximally filtered graph of bubble period (2007)



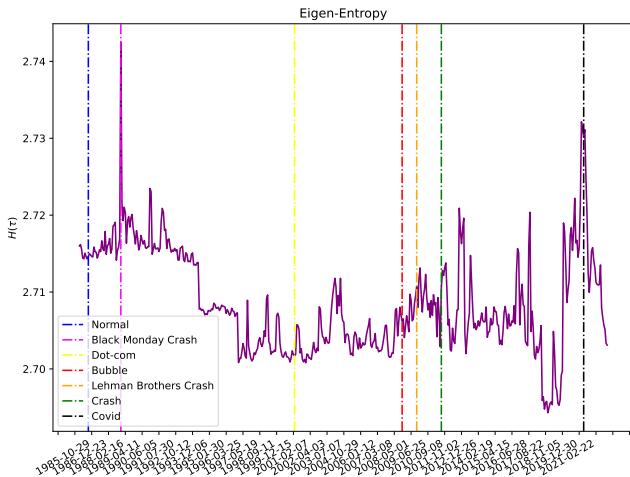
**Figure 48:** Planar maximally filtered graph of a normal market period (1985)



**Figure 50:** Planar maximally filtered graph of crash period (2010)



**Figure 51: Planar maximally filtered graph of covid crash (2020)**



**Figure 52: Temporal evolution of eigen-entropy**

Eigen-Entropy time series is another useful analysis: the higher the similarity among the stock centralities, the higher the eigen-entropy.

We define it as:  $H = - \sum_{i=1}^N p_i \log p_i$ , where  $p_i$  is the eigen-centrality of the  $i$ -th node (normalized).

From eigen-entropy time-series plot we can notice that the local maximum point correspond to crash period (as discussed before, during crash period the stocks are highly correlated and the stocks behave as if belonging to a single community).

## 8 DISCUSSION

In the previous section, we explored some possible means of investigating the current state of the stock market based

on its network characteristics. While we do not claim to have identified definitive criteria for defining the prevailing sentiment in the stock market, clearly impossible having only scratched the surface of the subject, we do feel we can point to directions for further research. In particular, there may be two refinements to be undertaken: on the one hand, a methodical study of the various historical series of the most important stock exchange indices, thus also of other countries and other epochs, in order to verify the extensibility of the impressions obtained here, and on the other hand, the construction of a more solid theoretical support that adapts the tools of social network science to the specific needs that the study of financial markets presents.

Studying the dynamics of the edges and how they change in being part of the PMFG could give us useful information about the strength/robustness and persistence of the graph itself.

Also using Machine Learning tool to find the best value for the threshold at different time epoch is another possible development.

## REFERENCES

- [1] R.N. Mantegna. 1999. Information and hierarchical structure in financial markets. *Computer Physics Communications* 121-122 (1999), 153–156. [https://doi.org/10.1016/S0010-4655\(99\)00302-1](https://doi.org/10.1016/S0010-4655(99)00302-1) Proceedings of the Europhysics Conference on Computational Physics CCP 1998.
- [2] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. 2005. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences* 102, 30 (2005), 10421–10426. <https://doi.org/10.1073/pnas.0500298102> arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.0500298102>