ANALYZING CLUSTER BEHAVIOUR AND PERFORMANCE IN PORTFOLIO RESEARCH

Mauro Leidi, Kilian Lock

Autumn semester 2021

1 Introduction

Clustering Methods are an old field. They were first established in 1932 to identify cultural relationship (1), but it took over 70 years to discover it's potential in finance. Finally, it was the semantic paper of Rosario Mantegna (2), fuelled by the rapid increase of computational power which opened cluster analysis for economic purposes. More than twenty years later, lots of different Clustering techniques have been developed (3). In this paper we will focus on the Louvain clustering (4) (5) and a maximum likelihood based clustering method introduced by Giada and Marsili (6). First we will have a look on basic properties of the algorithms such as computational costs and cluster creations as a function of parameters and assets, then we will proceed to check their stability and try to find an economic interpretation of the clusters.
In the second part of the paper, we want to investigate whether the specific clustering techniques will decrease the instability of the markowitz portfolio, as suggested by the research of Zhiwei Ren (7) and in a more general setting see, whether we can boost returns if we use the techniques on other trading strategies such as risk parity and equally weighted portfolios.

2 Methodology

In our paper we will use two clustering methods called Louvain and MLE. We will then proceed to check their properties and will see whether there is some economic sense behind the clusters. Finally we will check whether the clusters boost performance by using different strategies. The data which we will use for our research are random stocks across the globe and their respective close-to-close log-returns.

2.1 Clustering Methods

The clustering methods which we will use have been implemented by Mac Mahon & Garlaschelli (5) and Giada & Marsili (6), called Louvain method and MLE method respectively.

2.1.1 Louvain

The Louvain clustering is a community detection algorithm. It starts with a maximum entropy model, the null model, and builds a network on the base of it. Very quickly it gets computationally expensive. In our case the algorithm looks like following:

1. At the beginning we have a matrix of different assets and their respective returns. The first step is to create this maximum entropy model. In our case we started to correlate the returns, calculated its eigenvectors and eigenvalues to calculate recursively following formula until a threshold is achieved:

$$C_{null} = C + eigenvalue * eigenvector * eigenvetor^T$$

2. Now one needs to check each node independently. Each other asset needs to be checked whether there is an improvement if one adds $asset_i$ to the cluster j. If following increases, we add this $asset_i$ to the cluster j.

$$Q(\{asset_i\}) = \frac{1}{C_{tot}} \sum_{ij} [C_{i,j} - C_{null}]$$

3. repeat the second point for all nodes until a threshold is reached.

2.1.2 MLE

In our code we used a MLE method which is based on a merging algorithm. At the core of the idea lies a hierarchical clustering, the clusters which are formed by this method should have a similar cross-correlation. On a more technical side, in this method we try to maximize L_c. In our case the likelihood function is:

$$L_c = \frac{1}{2} \sum_{s,n_s} \left[ln\left(\frac{n_s}{c_s}\right) + (n_s - 1)\, ln\left(\frac{n_s^2 - n_s}{n_s^2 - c_s}\right) \right]$$

where s is a class and s_i is a specific asset. The definition of n_s and c_s is accordingly:

$$n_s = \sum_i \delta_{s,s_i}$$

$$c_s = \sum_{i,j} \delta_{s,s_i} \delta_{s,s_j} C_{i,j} = \sum_{i,j} \delta_{s,s_i} \delta_{s,s_j} Corr[s_i, s_j]$$

As explained before we will then use the merging algorithm. This works as following:

1. We start with N clusters, the amount of assets which we have in our dataset.
2. Then we start to merge those N clusters step by step such that the cost L_c of our new clusters is minimal.
3. Repeat step 2 N-1 one times so that in the end we have one single cluster.

2.2 Cluster Analysis

In the second section of the report we will focus on three different subcategories.

- Properties - We will have a look at computational costs and the amount of clusters in different settings. We will play with the amount of datapoints and assets to see fundamental behavior of the algorithms.
- Stability - We will have an in detail look in the different assets and see whether they change their cluster if one repeats the experiment. The idea is then to sort them accordingly by a n-pass clustering algorithm to see underlying structures.
- Sectors - Finally we will have a look whether we can talk some economic sense into our analysis by checking the industry, the stock exchange and the market capitalization of our different clusters.

2.3 Performance

We will have a look at three different portfolios. The Markowitz, Risk Parity and an equally weighted portfolio. We will do those performance measurements on inter and intra cluster portfolios a discretion.

2.3.1 Markowitz Portfolio

The Markowitz method was published in order to create the best possible trade-off between risk, as measured in volatility, and return, hence it boosts the sharpe ratio to a maximum if the estimated return μ and the estimated covariance Σ is correct.
To get the Markowitz tangency portfolio we create the following mathematical problem, w is the respective weights vector:

$$min\left(\frac{1}{2}w^T\Sigma w\right) \quad s.t.\ \ \mu^T w = \mu_p\ ;\ \ \mathbb{1}^T w = 1$$

We will solve this optimization problem with the Lagrangean:

$$\mathcal{L} = \frac{1}{2}w^T\Sigma w + \gamma(\mu_p - \mu^T w)$$

$$\frac{\partial \mathcal{L}}{\partial w} = \Sigma w - \gamma\mu$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \mu_p - \mu^T w$$

Solving this for w leads us to

$$w = \gamma\Sigma^{-1}\mu$$

Now we use the fact that $\mathbb{1}^T w = 1$ which leads us to

$$\gamma = \frac{1}{\mathbb{1}\Sigma^-1\mu}$$

and finally after inserting this term back we get:

$$w = \frac{\Sigma^{-1}\mu}{\mathbb{1}^T\Sigma^-1\mu}$$

As a somewhat more stable alternative to the general Markowitz Portfolio we will add a possibility to ban short selling, the only thing which changes is that now we add the condition that $w_i \geq 0$; $\forall i$. A third variant allows to set the maximum value of single weights, adding the contraint $w_i \leq C$; $\forall i$.

2.3.2 Risk Parity

One of the first risk parity portfolios was introduced in 1996 by a hedge fund of Bridgewater Associates and academically established by PanAgora Asset Management (8). The idea of this kind of portfolios is to boost the returns by increasing the weight of low volatility stocks. Weights are therefore determined by the following formula.

$$w_j = \frac{1}{\sigma_j \sum_{i=1}^{n} \sigma_i^{-1}}$$

We will be fully invested in risky assets in this strategy.

3 Results

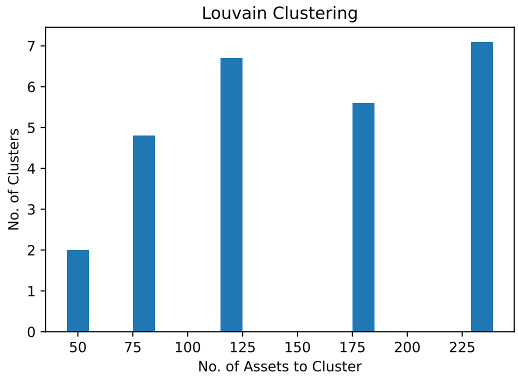
In this part of the report we are going to have an in detail look of the created Clusters and how their performance behave if used in portfolio analysis.

3.1 Cluster Techniques - An Overview

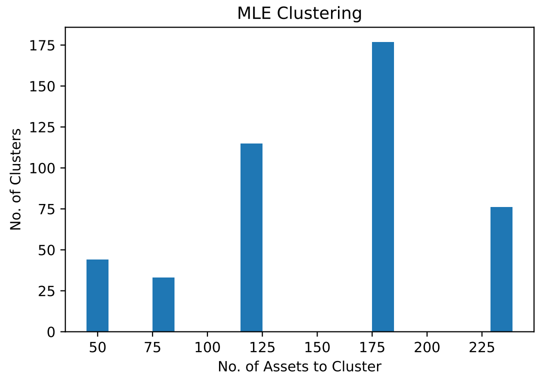
Following experiments have been made, if not stated differently, by averaging the results of ten runs. The x-scale is a real scale, though the numbers tested have been taken randomly.

3.1.1 Number of Clusters

In the first part of the research we wanted to see how the different clustering methods work and behave. In order to understand fundamental behavior we therefore plotted the amount of clusters which are created, depending on the amount of assets we gave. The results can be seen in figure 1 for the two clustering algorithms respectively. It depicts the most stable amount of clusters. We want to point out that in the MLE clustering, the amount of classes stays always the same, hence contrary to the Louvain method one does not need to redo the experiment several times. The number of Assets which we checked was picked randomly. What one can immediately see, the MLE clustering creates by far more classes as it seems to stabilize very quickly. A very strange behavior is that it increases if we decrease the amount of assets. This is a surprising pattern as we always took the assets from the same pot, hence their relation seem to be less strong in some settings.



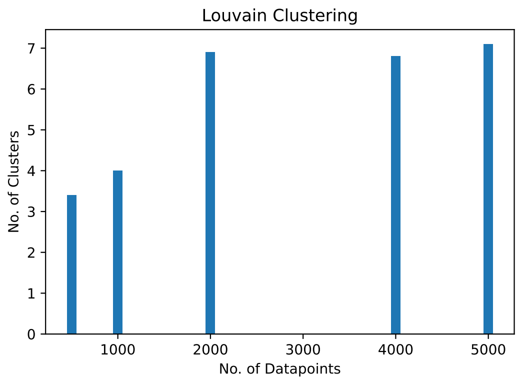
(a) Louvain Method



(b) MLE method

Figure 1: Number of clusters created in dependence of the number of assets available

Now on a less surprising side, the amount of clusters also changes if we reduce the number of datapoints (returns), which are in our case the parameters. This can be nicely seen in figure 2. Especially in the MLE clustering the amount of clusters when we use 4000 datapoints seem to be very small.



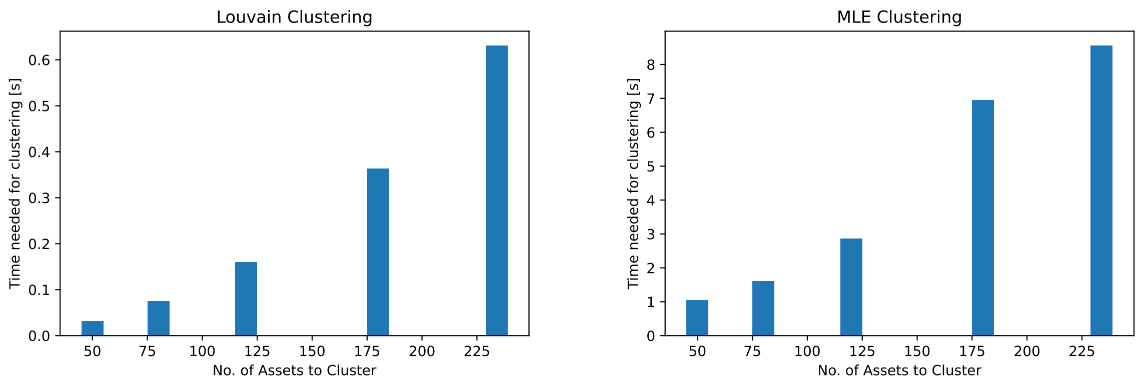
(a) Louvain Method

(b) MLE Method

Figure 2: Number of clusters created in dependence of the number of datapoints (returns)

3.1.2 Computational Cost

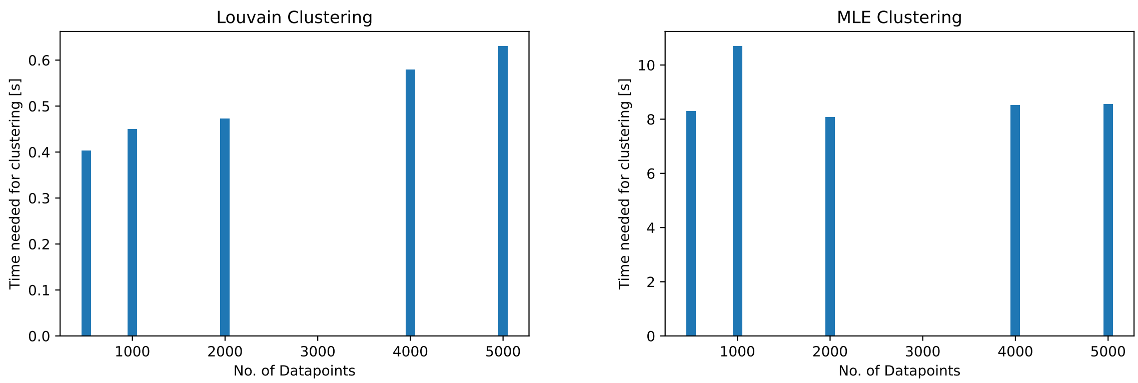
Secondly, when looking at the two techniques, we wanted to see how much time they needed to cluster the data. The hypothesis is that the computational time is dependent on two things: amount of assets and amount of datapoints.
As one can see in figure 4, clearly the computational time decreases exponentially when decreasing the number of assets in a cluster. This is as expected. On a more surprising side, even though Louvain is a network method it seems to be more than ten times faster in computation than its MLE counterpart. Though this might be due to a not optimized algorithm or to the fact that we are using an R package in python using rpy2 library as bridge between the two programming languages.



(a) Louvain Method (b) MLE Method

Figure 3: Computational time needed to cluster in dependence of the number of assets available

If one looks at figure 4, which depicts calculation time as a function of parameters of a single cluster we see some unsuspected behavior. For the Louvain clustering, hence figure 5a, one can see the amount of time decreases nicely, though for figure 5b we observe some more randomness. What is especially strange is if one considers figure 2b, it does not seem to correlate with the cluster size at all.



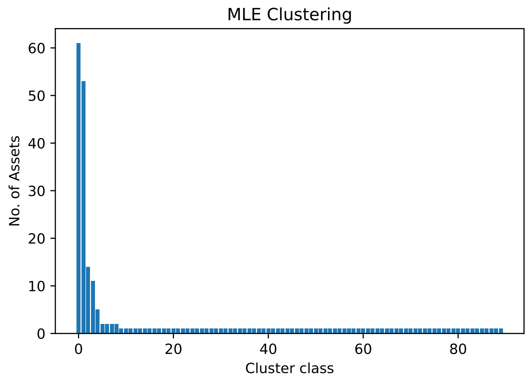
(a) Louvain Method (b) MLE Method

Figure 4: Computational time needed to cluster in dependence of the number of datapoints (returns)

3.1.3 Amount of Assets in Clusters

As a next step we wanted to see how the distribution of the different classes are in a specific environment and whether the clusters which are formed cover big chunks of the data or not. You can see in figure 5 that at least for the Louvain clustering, the differences are not that big. In contrast, the MLE clearly has some very big clusters and then exponentially decreasing smaller ones. From cluster 12 on, nearly all of them are a single asset where the algorithm could not detect any similarities.

(a) Louvain Method



(b) MLE Method

Figure 5: Cluster sizes for the two different models

Figure 6 is the reference for Figure 2b where we had the small amount of clusters. Again we see an exponential decrease.

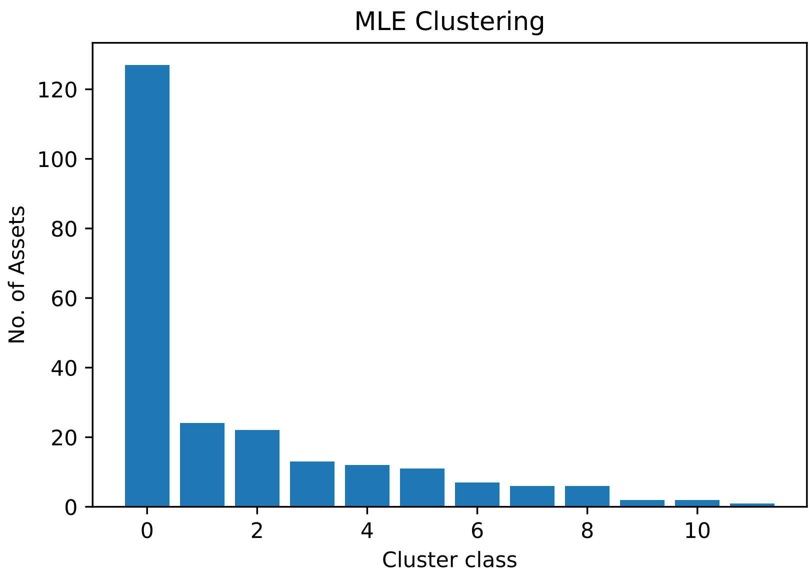


Figure 6: Cluster size in the specific setting where we reduced the datapoints from 5000 to 4000

3.2 Asset Observation in Specific Clusterings

As a second question we focused on how different clusters behave. That means, how stable is their classification and whether the grouping make sense from an economic perspective, hence investigating the clusters in regards to industry, stock exchange and market capitalization.

3.2.1 Stability

In the following section we will have an in-detail look into five different assets from different industries. Namely:

- GE - General Electric is a big US conglomerate. It includes sectors such as aviation, power and finance. The motivation to investigate it comes through its broad bussiness.
- XOM - Exxon Mobile is a oil and gas company and thus lays the foundation of a lot of industries. As a commodity company it should be more volatile and therefore create its own class with other oil companies.
- JNJ - Johnson & Johnson is a pharmaceutical company. As health is decoupled from recessions and booms one might think it will create its own bubble.
- MCD - Mc Donalds. Again a rather specific asset.
- S&P 500 - The largest companies are included in the S&P 500 index. It should cluster with a lot of assets.

When looking at stability we took the full dataset for the louvain and the MLE method. As explained before, the MLE method is stable and the plot will jump between zero and one only as you can see in figure 7. Stability in this case (MLE) won't be further discussed and we focus on Louvain.

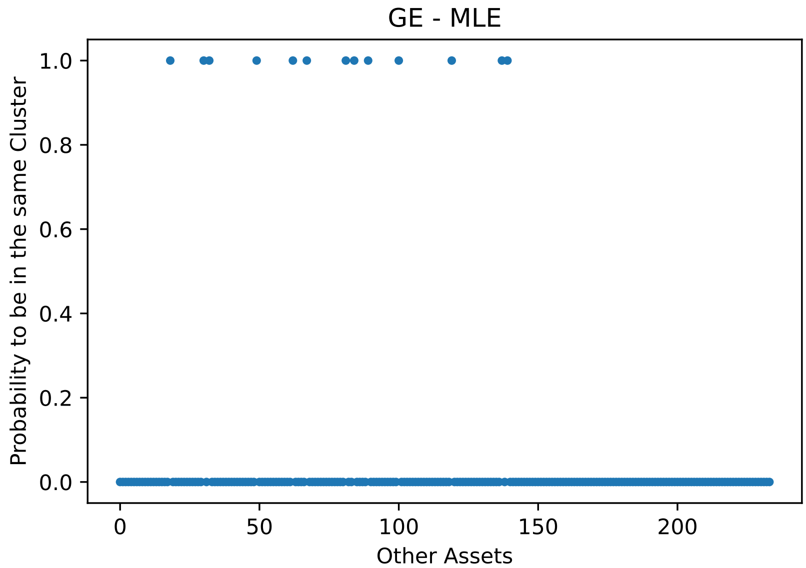


Figure 7: The cluster method by MLE gives always the same results if the same initial conditions appear.

Now if one looks at figure 8 one can see that GE seems to be rather unstable and likes to build different clusters compared to XOM, JNJ and MCD. Especially interesting is the S&P 500 which seems, overall, to include lots of different assets which is what we expected.

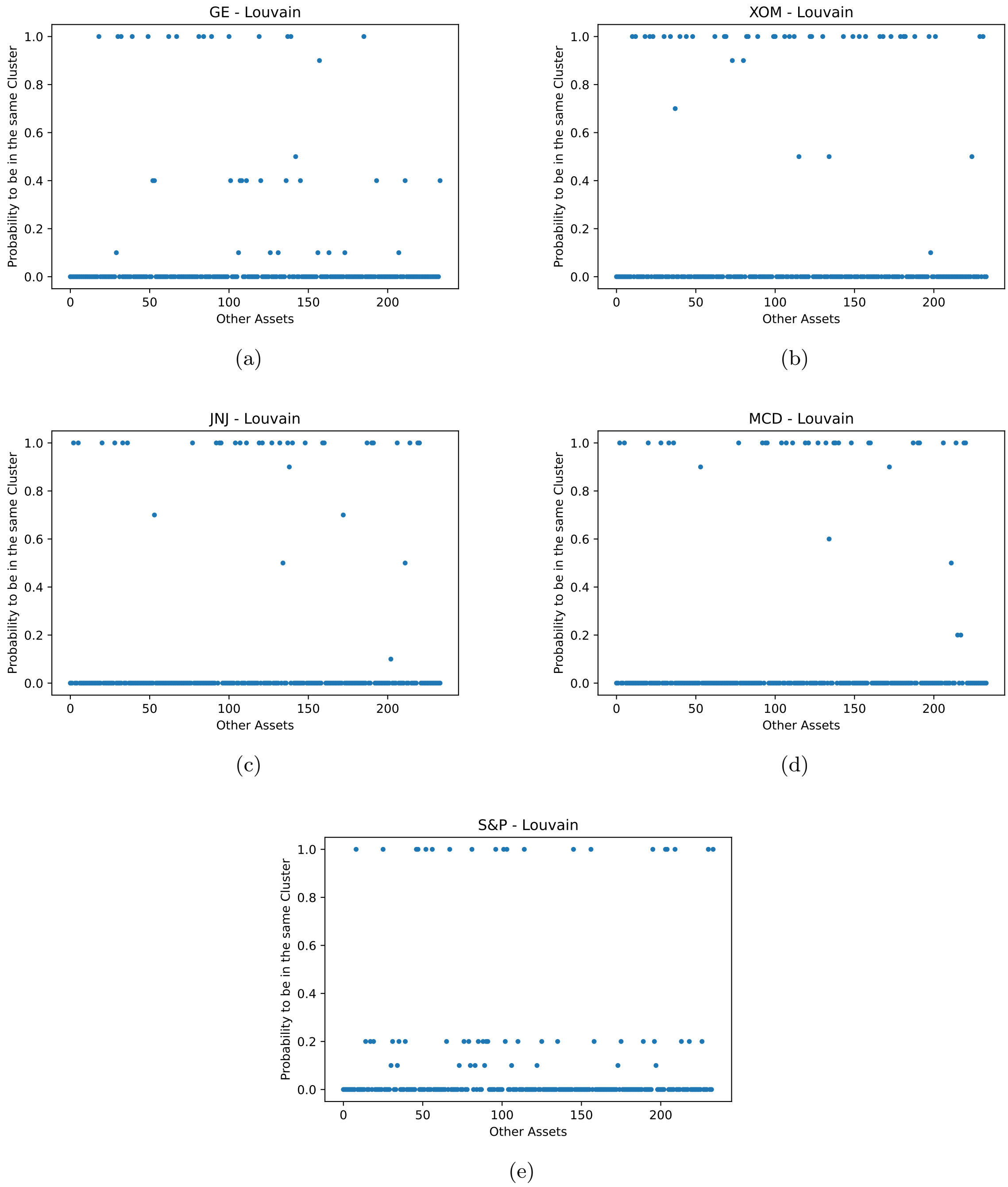


Figure 8: Louvain Clustering Stability for different specific assets

Now if one looks at this more broadly, one can have a look at figure 11. It is the correlation map for every individual asset. They are ordered in alphabetical order and one can have a look at them in the Appendix.

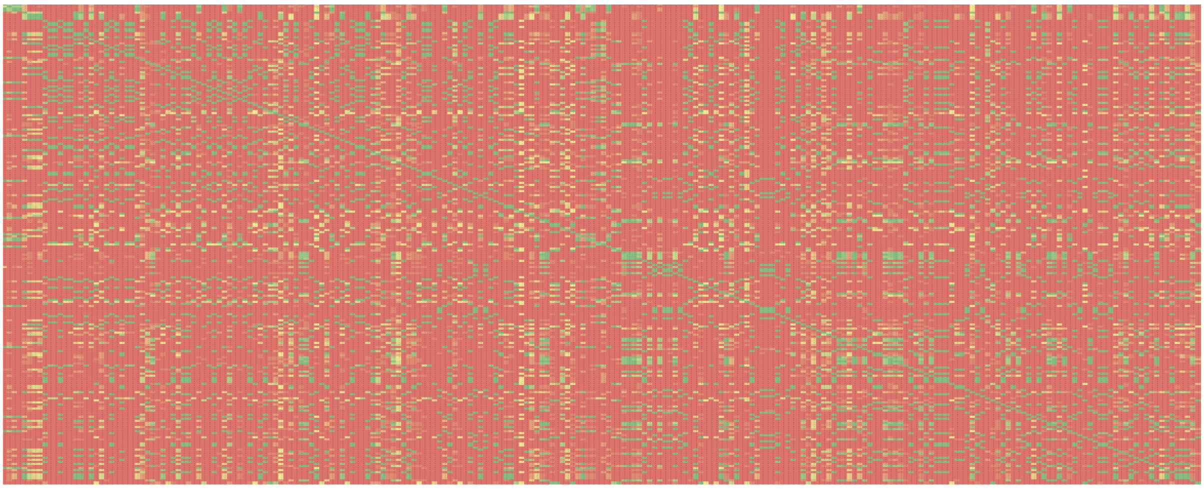


Figure 9: Louvain clustering method and the stability of the different asset. Red means they are never in the same cluster and green says they are always in the same cluster (10 runs).

By some n-pass clustering algorithm over the correlation matrix we are able to order this matrix a bit more. Below, in figure 10, is the same correlation matrix as before. As a small sidenote for the reader: The plot is a .pdf plot, meaning one can zoom in massively and still read the assets which are on the x and y axis.

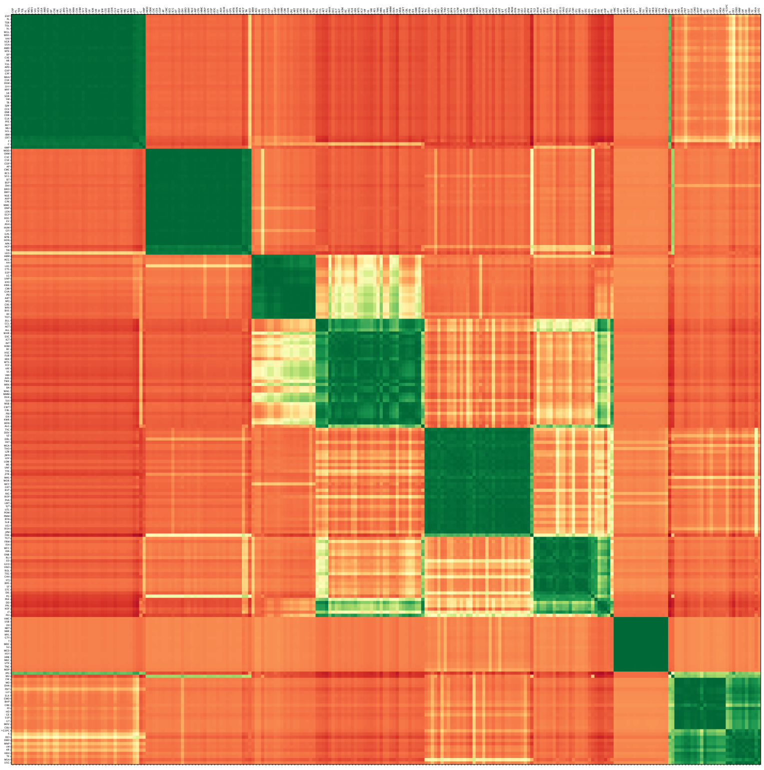


Figure 10: Louvain clustering method and the stability of the different asset. Red means they are never in the same cluster and green says they are always in the same cluster (10 runs).

If we repeat this experiment for the MLE clustering we get figure 7. As the algorithm is stable we had 1 run, thus only green and red cells. One can nicely recognize the diagonal this time.

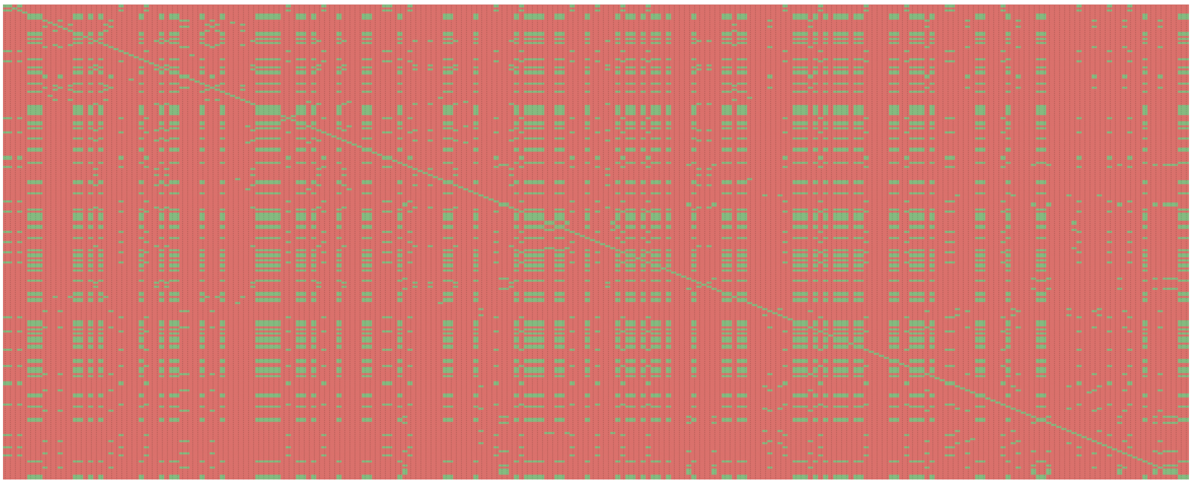


Figure 11: MLE clustering method and the stability of the different asset. Red means they are never in the same cluster and green says they are always in the same cluster (1 run as stable).

Again, when ordering we get the results presented in figure 12 and immediately spot the big amount of lone clusters.

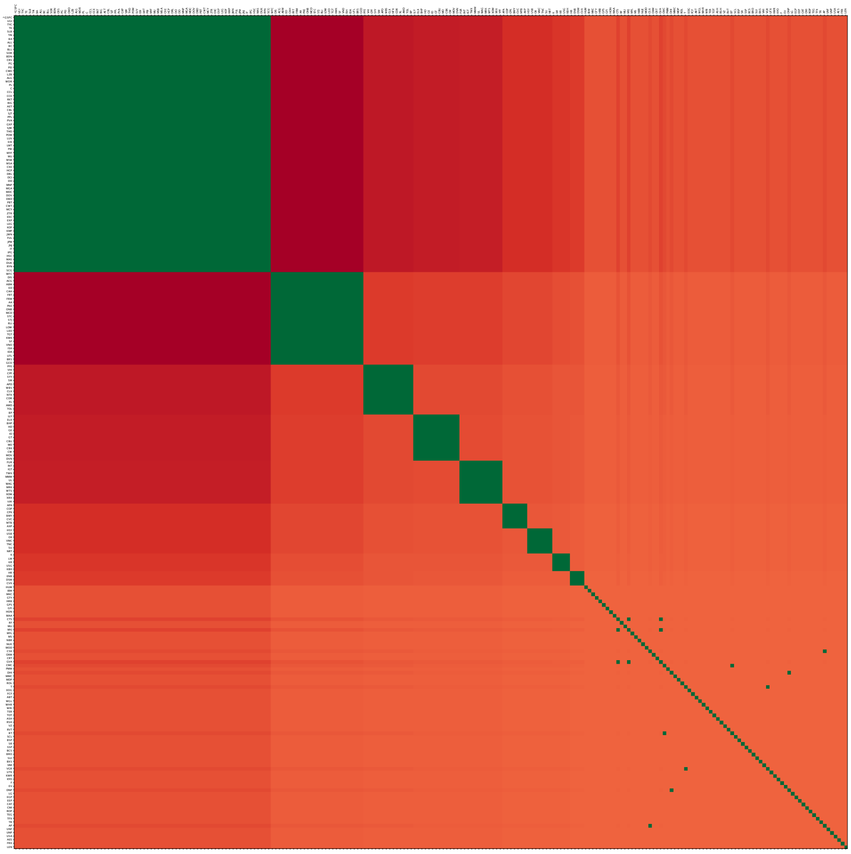


Figure 12: MLE clustering method and the stability of the different asset. Red means they are never in the same cluster and green says they are always in the same cluster (1 run as stable).

3.2.2 Comparison

In order to compare both methods we colored the clusters of the Louvain Method by clusters, tracked them down in the MLE method and gave them the same color. When one looks at figure 13 of the next page, one can quickly see they somehow, in the smaller clusters (Cluster 2, 3, 4, 5, 6, 7, 8, 9), coincide quite heavily. This is a very pleasing result and shows that the algorithms behave similarly. For further analysis we will thus focus on the Louvain Method. All the following plots can be found in the Appendix for the MLE method. Also, the order of the assets will always be the same for the following analysis.

Asset No.	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
1	AMD	AP	ABM	AES	CAT	BRS	GTY	^GSPC
2	APD	APA	ACG	AET	DOV	CAH	HRB	BHP
3	BP	ASH	AA	ALG	DRL	DD	HSY	CBK
4	C	AXP	LEG	ALL	FRT	DIS	MCD	CBU
5	CCE	BCS	CTS	BA	FRX	ED	MDC	CMO
6	CDR	BMY	EEP	BC	JPM	FRM	MDP	CW
7	CLX	BRO	LG	BDN	JWN	GCO	MKC	DBD
8	CPF	BSP	UNP	BLL	LMT	IDA	MYL	DCI
9	CRT	BT	KYO	CBL	LOW	ISH	NBR	DVN
10	CVX	CMC	KWK	CCL	LUV	IT	NRT	ELX
11	CXE	COP	CMI	CWT	LZB	KOF	OII	GE
12	DSM	CPK	GVA	DUK	MAS	ONB	SU	GT
13	ENB	CSX	JNJ	EIX	MCY	PEI	T	HD
14	EXP	CVC	ABT	EXC	MGA	PG	TNC	IO
15	F	DHI	MYJ	FCF	MS	PKE	USB	IPG
16	FUL	DNP	CRS	FUR	PMM	PPL	UTX	K
17	GXP	EGP	WRI	HSC	POM	RLI	VMC	KBH
18	HR	EOG	BXS	IGT	PVA	ROL		KR
19	IBM	EV	RF	INT	RIG	SF		LLY
20	KMP	GFI	TOT	KWR	ROP	STC		LM
21	MAA	GPS		MMM	RSH	STJ		MNP
22	MLI	HCP		MRK	RVT	TEG		MO
23	MYF	HON		MSB	RYN	TGT		MOV
24	NTX	HUM		PBI	SJT	VNO		MSA
25	PFE	LEN		SUI	SLB	WFC		MU
26	PL	LEO		TWX	TFX			PBT
27	RKT	MMC		UL	THO			SSP
28	SCL	MOD		VMI	TYC			TE
29	SJW	MTB		WAG	UGI			USG
30	SM	NUE		WHR	UNF			
31	SOR	OXM		WTS	UTL			
32	SR	SCG		XOM	VZ			
33	SYY	TAI		XRX	WOR			
34	TER			YPF	ZTR			
35	TOL							
36	TR							
37	USA							
38	VGR							
39	VHI							
40	WBS							
41	WGL							
42	XL							

Asset No.	Cluster 1		Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1	^GSPC	MDC	AA	AMD	BHP	FUR	APA	HSY	K	CVX
2	AET	MGA	ABM	APD	CBK	IGT	AXP	NRT	KBH	DSM
3	ALG	MNP	ACG	BP	CBU	INT	BMY	OII	KR	ENB
4	ALL	MSA	BRS	CDR	CW	MMM	COP	SU	LM	HR
5	BA	MSB	CAH	CLX	DVN	MRK	CPK	TNC	USG	
6	BC	MU	DIS	CPF	ELX	TWX	CVC	USB		
7	BDN	MYF	ED	NTX	GE	UL	MTB	VMC		
8	BLL	PBI	FRM	PFE	GT	WAG				
9	C	PBT	FRT	SM	HD	WTS				
10	CBL	PEI	GCO	SYY	IO	XOM				
11	CCE	PG	IDA	TOL	LLY	XRX				
12	CCL	PL	ISH	VHI	MO	YPF				
13	CMO	POM	KWK	WBS	MOV					
14	CRS	PPL	LEO	XL						
15	CWT	PVA	LOW							
16	CXE	RIG	MCD							
17	DBD	RKT	ONB							
18	DCI	RYN	PKE							
19	DD	SCG	RLI							
20	DOV	SJT	SF							
21	DRL	SJW	STC							
22	DUK	SLB	STJ							
23	EIX	SOR	TGT							
24	EXC	TAI	UTL							
25	EXP	TE	VNO							
26	FUL	THO	WFC							
27	GXP	TYC								
28	HCP	UGI								
29	HSC	WOR								
30	IPG	ZTR								
31	IT									
32	JNJ									
33	JPM									
34	JWN									
35	KMP									
36	KOF									
37	LEG									
38	LMT									
39	LUV									
40	LZB									
41	MAS									
42	MCY									

Figure 13: Comparing MLE (lower plot) and Louvain (upper plot).

3.2.3 Industry

The first thing which we were wondering is whether the clusters are built on the base of industry sectors. We selected a few and colored them which can be seen in figure 14. What we notice is that there seems to be no structure here, the sectors are very scattered in the different clusters.

Asset No.	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
1	Technology	Mining	Consumer Goods	Technology	Machinery	Mining	Real Estate	Index
2	Chemicals	Oil	Pharmaceuticals	Oil	Machinery	Food	Consumer Goods	Mining
3	Oil	Chemicals	Mining	Hotels	Oil	Chemicals	Food	Banking
4	Banking	Banking	Consumer Goods	Hotels	Real Estate	Media	Hotels	Banking
5	Real Estate	Banking	Technology	Aerospace	Telecommunications	Technology	Homebuilding	Consumer Goods
6	Consumer goods	Pharmaceuticals	Healthcare	Leisure	Banking	Food	Real Estate	Aerospace
7	Consumer Goods	Insurance	Technology	Real Estate	Consumer Goods	Consumer goods	Food	Technology
8	Banking	Index	Freight	Containers & Packaging	Aerospace	Technology	Mining	Real Estate
9	Index	Hotels	Technology	Real Estate	Consumer Goods	Pharmaceuticals	Oil	Oil
10	Oil	Mining	Automobiles	Hotels	Freight	Software	Oil	Technology
11	Index	Oil	Automobiles	Water	Consumer Goods	Beverages	Oil	Consumer Goods
12	Chemicals	Oil	Construction	Technology	Homebuilding	Banking	Oil	Automobiles
13	Oil	Freight	Pharmaceuticals	Technology	Insurance	Real Estate	Telecommunications	Consumer Goods
14	Construction	Banking	Healthcare	Technology	Automobiles	Consumer Goods	Machinery	Oil
15	Automobiles	Homebuilding	Index	Banking	Banking	Aerospace	Banking	Media
16	Hotels	Food	Mining	Oil	Banking	Multiline_Utilities	Aerospace	Food
17	Banking	Real Estate	Mining	Commercial Services	Automobiles	Insurance	Construction Materials	Homebuilding
18	Real Estate	Oil	Index	Hotels	Food	Consumer Goods		Food
19	Software	Mining	Banking	Oil	Oil	Banking		Pharmaceuticals
20	Paper	Mining	Oil	Chemicals	Software	Insurance		Aerospace
21	Real Estate	Consumer Goods		Consumer Goods	Healthcare	Banking		Index
22	Machinery	Technology		Pharmaceuticals	Banking	Real Estate		Food
23	Banking	Consumer Goods		Mining	Real Estate	Consumer Goods		Textiles
24	Software	Healthcare		Office_Equipment	Oil	Real Estate		Consumer Goods
25	Pharmaceuticals	Homebuilding		Real Estate	Oil	Banking		Hotels
26	Aerospace	Machinery		Media	Healthcare			Oil
27	Banking	Insurance		Personal Services	Consumer goods			Media
28	Chemicals	Automobiles		Construction	Banking			Telecommunications
29	Water	Banking		Real Estate	Oil			Mining
30	Oil	Mining		Consumer Goods	Commercial Services			
31	Banking	Textiles		Machinery	Consumer goods			
32	Oil	Real Estate		Oil	Telecommunications			
33	Food	Freight		Office_Equipment	Mining			
34	Technology			Oil	Index			
35	Homebuilding							
36	Food							
37	Index							
38	Food							
39	Chemicals							
40	Banking							
41	Mining							
42	Automobiles							

Figure 14: Does the Louvain clustering build industry patterns

3.2.4 Stock Exchange

Hence we continued our research onto the stock exchange. Though one sees immediately, when looking at figure 15, that our dataset was dominated by the New York Stock Exchange. The smaller exchanges are again very scattered throughout the clusters which hence make it hard to talk economic sense into it.

Asset No.	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
1	XNAS	XNYS	XNYS	XNYS	XNYS	XWAR	XNYS	XNYS
2	XNYS	XNAS	XCNQ	XLON	XNYS	XNYS	XNYS	XNYS
3	XLON	XNYS	XNYS	XASX	XLON	XNYS	XNYS	XFRA
4	XNYS	XNYS	XNYS	XASX	XNYS	XNYS	XNYS	XNYS
5	XWAR	XNYS	XNYS	XNYS	XASX	XNYS	XNYS	XNZE
6	XWAR	XNYS	BMEX	XNYS	XNYS	XASX	XWAR	XNYS
7	XNYS	XNYS	XKRX	XNYS	XNYS	XNYS	XNYS	XNYS
8	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XASX	XLON
9	XNYS	XNAS	XTOK	XNYS	XNYS	XTSX	XNYS	XNYS
10	XNYS	XNYS	XTOK	XNYS	XNYS	XNYS	XNYS	XLON
11	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS
12	XAMS	XNYS	XNYS	XNYS	XNYS	XNAS	XNYS	XSWX
13	XNYS	XNAS	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS
14	XNYS	XASX	XNYS	XNAS	XNYS	XNYS	XNYS	XNYS
15	XSWX	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS
16	XLON	XWAR	XNYS	XAMS	XNYS	XNYS	XSWX	XNYS
17	XFRA	XNYS	XTSX	XNYS	XPAR	XNYS	XNYS	XNYS
18	XNYS	XNYS	XNYS	XNYS	BMEX	XNYS		XNYS
19	XNYS	XTSX	XNYS	XNYS	XNYS	XNYS		XNYS
20	XWAR	XNYS	XTSE	XNYS	XNYS	XNYS		XNYS
21	XNYS	XNYS		XNYS	XASX	XLON		XNYS
22	XNYS	XNYS		XNYS	XNYS	XFRA		XNYS
23	XNYS	XNAS		XNYS	XNYS	XNYS		XNYS
24	BMEX	XNYS		XNYS	XNYS	XNYS		XNYS
25	XNYS	XNYS		XNYS	XNYS	XNYS		XNYS
26	XNYS	XETR		XNYS	XNYS			XNYS
27	XNYS	XNYS		XNYS	XNYS			XNAS
28	XNYS	XNYS		XNYS	OTCM			XLON
29	XNYS	XNYS		XFRA	XNYS			XNZE
30	XNYS	XNYS		XNYS	XNYS			
31	XNYS	XNYS		XNYS	XNYS			
32	XNYS	XASX		XNYS	XNYS			
33	XNYS	OTCM		XNAS	XNYS			
34	XNAS			XNYS	XNYS			
35	XNYS							
36	XNYS							
37	XNYS							
38	XNYS							
39	XNYS							
40	XNYS							
41	XNYS							
42	XNYS							

Figure 15: Stock exchanges of the Louvain Method.

3.2.5 Market Capitalization

Finally, we had a look at whether the method clusters in regards to market capitalization and plotted the results in figure 16. Finally, also in this plot it is difficult to find a pattern on the base of market cap. For us, this really means that the algorithms finds a new, unexpected way to cluster and that neither market cap, nor stock exchange or industry is incorporated in those clusters. This shows us that there are cross-similarities away from classical economic understanding.

Asset No.	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
1	1.7E+11	9.5E+07	2.7E+09	1.6E+10	1.1E+11	2.2E+08	1.5E+09	0.0E+00
2	6.7E+10	9.8E+09	8.5E+05	4.4E+07	2.6E+10	1.5E+10	4.2E+09	1.5E+11
3	8.8E+10	6.1E+09	1.1E+10	6.5E+08	1.7E+09	4.2E+10	4.0E+10	9.5E+09
4	1.2E+11	1.3E+11	5.5E+09	2.9E+10	1.1E+10	2.8E+11	2.0E+11	4.0E+09
5	8.0E+06	4.2E+10	1.2E+09	1.2E+11	2.4E+07	3.0E+10	4.4E+09	3.5E+08
6	4.9E+09	1.4E+11	6.4E+06	7.8E+09	4.7E+11	1.5E+07	1.3E+07	5.4E+09
7	2.1E+10	2.0E+10	1.1E+10	2.3E+09	3.6E+09	9.7E+08	2.6E+10	7.1E+08
8	7.9E+08	6.7E+08	1.6E+11	3.1E+10	9.8E+10	5.7E+09	1.3E+08	5.1E+07
9	7.6E+07	1.9E+07	2.2E+10	6.2E+08	1.7E+11	2.9E+08	6.7E+08	3.0E+10
10	2.3E+11	4.4E+09	6.1E+09	2.3E+10	2.5E+10	2.7E+10	1.0E+08	6.7E+10
11	1.7E+08	9.5E+10	3.1E+10	3.8E+09	1.6E+09	2.9E+09	1.1E+09	1.0E+11
12	3.9E+10	2.6E+09	1.8E+09	8.1E+10	1.7E+10	3.0E+09	4.6E+10	5.7E+09
13	1.0E+11	8.3E+10	4.5E+11	2.6E+10	2.9E+09	8.2E+07	1.8E+11	4.3E+11
14	6.8E+09	3.0E+08	2.5E+11	5.7E+10	3.1E+10	4.0E+11	1.5E+09	2.6E+07
15	4.8E+10	3.9E+10	3.9E+08	1.5E+09	1.8E+11	2.7E+08	8.3E+10	1.5E+10
16	1.6E+08	9.0E+09	1.4E+09	8.1E+08	4.1E+08	2.3E+10	1.2E+11	2.2E+10
17	8.2E+07	9.3E+09	1.3E+06	1.3E+09	3.8E+09	5.1E+09	2.8E+10	3.9E+09
18	4.7E+09	5.2E+10	5.1E+09	5.9E+09	1.4E+07	1.7E+10		3.4E+10
19	1.2E+11	1.4E+07	2.1E+10	1.7E+09	1.8E+09	7.4E+09		2.6E+11
20	2.8E+07	9.7E+09	2.7E+08	4.1E+09	5.2E+10	2.1E+09		9.8E+10
21	2.6E+10	6.6E+09		1.0E+11	4.6E+07	1.2E+10		1.7E+08
22	3.4E+09	1.3E+10		1.9E+11	2.1E+09	4.1E+09		8.7E+10
23	7.1E+08	1.4E+11		3.5E+08	5.8E+09	1.1E+11		9.6E+08
24	3.7E+07	6.0E+10		1.2E+09	3.2E+08	8.0E+09		5.9E+09
25	3.3E+11	3.5E+10		2.4E+10	4.2E+10	1.9E+11		2.3E+09
26	1.4E+09	3.7E+08		6.9E+10	1.5E+10			5.5E+08
27	2.8E+10	8.8E+10		1.4E+11	5.8E+09			1.6E+09
28	2.8E+09	5.2E+08		5.3E+09	1.1E+08			2.9E+10
29	2.2E+09	2.0E+10		1.1E+07	9.6E+09			2.2E+08
30	3.6E+09	3.3E+10		1.5E+10	4.0E+09			
31	4.0E+08	1.7E+09		6.5E+09	7.3E+08			
32	3.4E+09	1.6E+10		2.6E+11	2.2E+11			
33	4.0E+10	2.2E+08		4.0E+09	2.8E+09			
34	2.7E+10			3.0E+09	4.3E+08			
35	8.7E+09							
36	2.3E+09							
37	1.9E+09							
38	1.8E+09							
39	8.1E+08							
40	5.1E+09							
41	4.9E+06							
42	4.6E+08							

Figure 16: Comparing MLE and Louvain

3.3 Impact of Portfolio Strategies on Portfolio Performance and Risk

The intent of this section is to study the performance of clustering methods by means of a simple example and to analyze whether it is possible to outperform Markowitz's strategies. Past data will be used as expected returns, and we will analyze four different strategies: the two clustering techniques, and two variations of Markowitz's strategies.
It is clear that, if we want to estimate the performance of our strategy using returns that were already used to calculate the optimal portfolio, our performance estimator will be biased. A model that over-fits on past data would perform better than a model that somehow manages to predict the true probability distribution we are looking for, i.e. to select the portfolio that in expectation has the best future performance. In addition, we are already assured that this biased approach would show that the Markowitz's portfolio has better performances, because it is calculated in such a way that it maximizes its performance on the data which we use as expected returns. Precisely for this reason, the Markowitz's portfolio tends to overfit on the expected return we forecasted, and therefore has the highest sensitivity with respect to variations in the expected returns.
More formally we can understand it in a simple case of two stocks: given the covariance matrix

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} \\ c_{2,1} & c_{2,2} \end{pmatrix}$$

and a risk adversity λ we can find the optimal weights following the Markowitz's optimiization problem to be:

$$w_1 = \frac{\lambda * (\mu_1 * c_{2,2} + \mu_2 * c_{1,2})}{c_{1,1} * c_{2,2} - c_{1,2} * c_{2,1}}$$

$$w_2 = \frac{\lambda * (\mu_1 * c_{1,1} + \mu_2 * c_{2,1})}{c_{2,2} * c_{1,1} - c_{2,1} * c_{1,2}}$$

By introducing the correlation as

$$\rho_{1,2} = \frac{c_{1,2}}{c_{1,1} * c_{2,2}}$$

and defining the sensitivity as the partials derivative of the weights with respect to variations in the expected returns:

$$S_{1,1} = \frac{\partial w_1}{\partial \mu_1} = \frac{\lambda}{c_{1,1} * (1 - \rho_{1,2}^2)}$$

This metric allows us to estimate how much a difference between our estimate of future returns will have an impact on the weights of the portfolio and therefore also on the performance of the selected portfolio. To minimize the sensitivity of our portfolio to changes in expected returns, we should minimize the correlation term $\rho_{1,2}$. This motivates the clustering techniques that are achieving high inter-group correlation and low intra-group correlation.
For this reason clustering strategies are less sensitive to changes in expected returns. Despite this, we know that if our expected returns are correct, their performance is not optimal, and therefore doubts persist as to which approach is better. To answer this question, we divided the returns into intervals of N days, and used each interval as a forecast of the expected returns of the next interval. We then used these expected returns to calculate the optimal portfolio given a strategy, and calculated the realized returns from the next interval. The result is equivalent to the simple strategy whereby the composition

of the portfolio is changed every N days, using only the returns of the N previous days as a metric to compute expected returns. To keep things simple we used equal weights for combining stocks inside a cluster.
To have greater statistical robustness in our results we have simulated this strategy many times, and each time we have generated a different dataset by sampling at random K stocks, and then using only the returns of the selected stocks to perform a single experiment. We repeated the experiment 1500 times with K = 75,45,20 stocks and N = 90 days. Additionally, to measure whether weight-constraining strategies have an influence on sensitivity, helping to avoid overfitting, and to quantify their effect, we tested two different max sharpe rate portfolios: the first free to take weights between [-1,1] the second can only assume values between [-0.1,0.1]. We will name the first method Markowitz's free weight method and the second one Markovitz WL.

Sharpe Rateos	Louvian	GiadaMarsili	Markovitz	Markovitz WL
75 stocks	0.611	0.637	0.579	0.576
45 stocks	0.616	0.627	0.539	0.616
20 stocks	0.599	0.598	0.466	0.656

Table 1: Sharpe rateos for different values of K and different strategies.

P values K= 75	Louvian	GiadaMarsili	Markovitz	Markovitz WL
Louvian		0.0217	9.8e-8	6.28e-10
GiadaMarsili			0.00134	4.41e-5
Markovitz				0.39

Table 2: P values of t-test for K = 75 stocks.

As we can see in figure 17, for K = 75 stocks, Markowitz's methods tend to have a higher mean return, but at the same time, being more sensitive they also have a higher volatitlity. We perfomed a t-test to test for the difference in means and found different p values presented in table 2. The difference between the log returns of the two Markowitz's methods is not statistically significant. The difference in means between pairs of (clustering method, Markowitz's method) is always statistically significant. In table 1 we can observe sharpe rateos: we can see that Louvian Clustering is performing best, followed by GiadaMarsili and finally by Markowitz's methods. clustering technique are able to outperform Markowitz's methods using this simple strategy. This shows the importance of having a method that is less sensitive to our prediction of future returns.

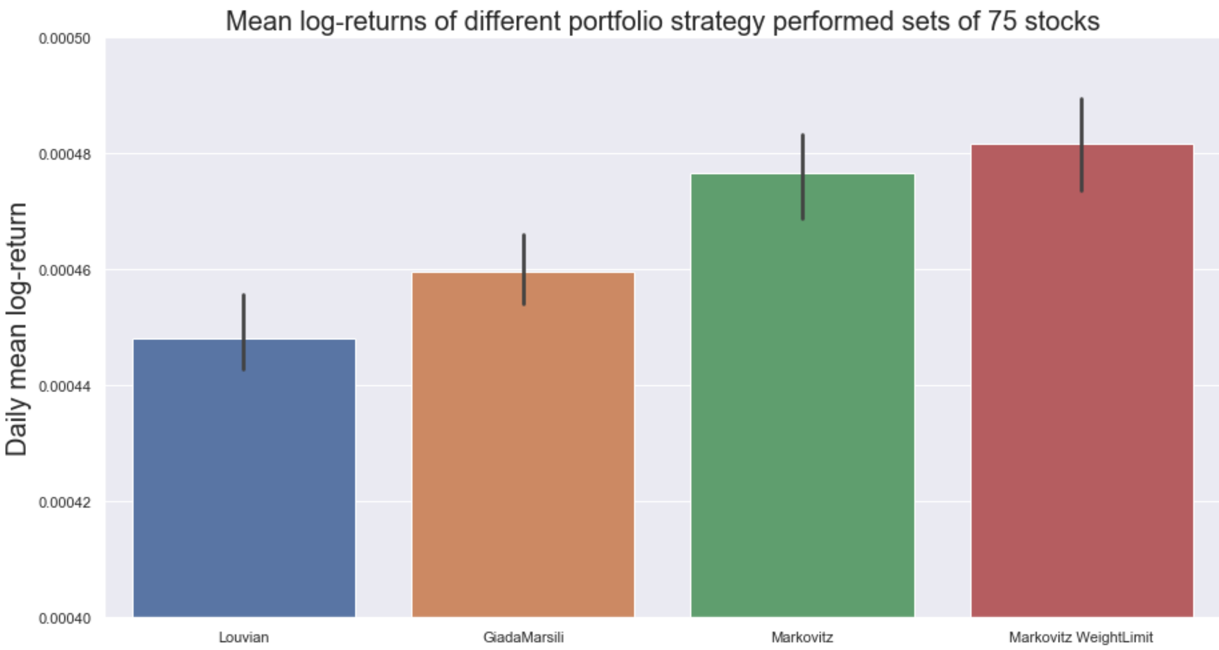


Figure 17: Mean log returns with different portfolio strategies on sets of 75 stocks. Markowitz's optimization leads to higher mean daily retrurns. The errorbars shows 95% confidence intervals computed with a bootstrapping technique.

P values K = 45	Louvian	GiadaMarsili	Markovitz	Markovitz WL
Louvian		0.39	0.912	1.45e-07
GiadaMarsili			0.366	7.64e-10
Markovitz				1.53e-06

Table 3: P values of t-test for K = 45 stocks.

As we can see in figure 18, for K = 40 stocks, Markowitz's free weights method tend to have a comparable mean log return to clustering techniques, but at the same time, being more sensitive has a higher volatitlity. On the other hand Markowitz's WL method has the highest mean log return. As we can see from table 3 the difference in means between pairs of methods is not statistically significant except for pairs containing the Markowitz's WL method. In table 1 we can observe sharpe rateos: clustering technique are able to outperform Markowitz's methods using this simple strategy. It's important to note that the Markowitz's WL method performs almost as good as clustering techniques and way better than the Markowitz's free weight method.

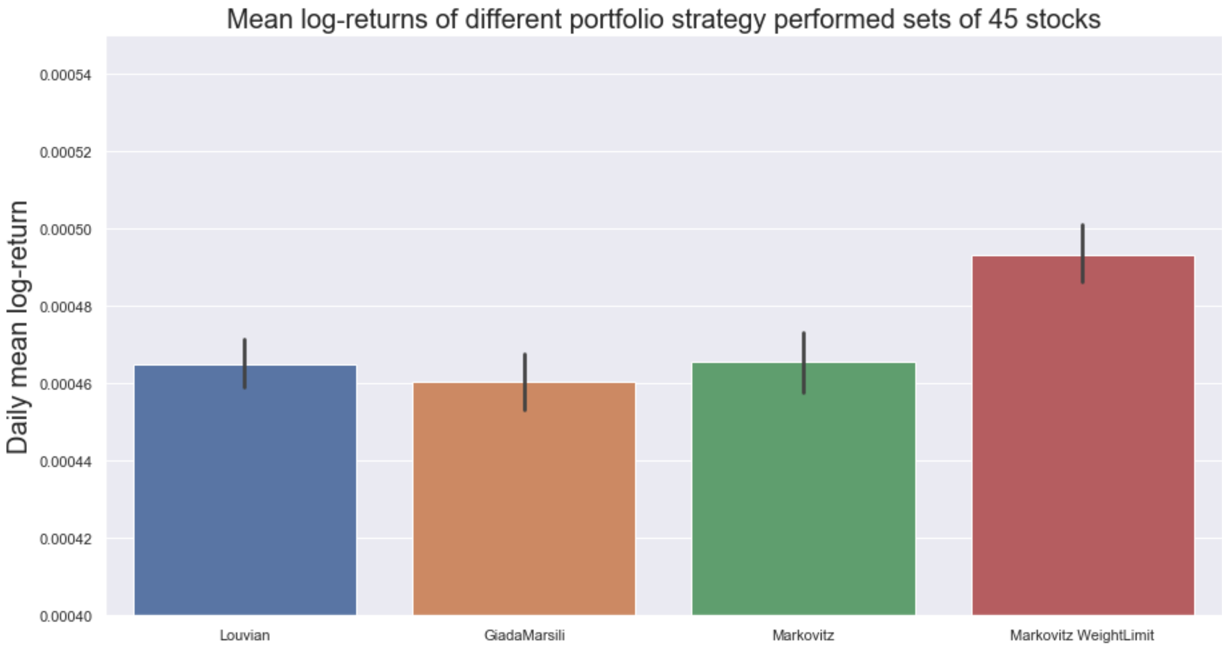


Figure 18: Mean log returns with different portfolio strategies on sets of 45 stocks. Markowitz's free weights optimization and Clustering techniques yield similar mean returns. Markowitz's WL method has a higher mean return. The errorbars shows 95% confidence intervals computed with a bootstrapping technique.

P values K = 20	Louvian	GiadaMarsili	Markovitz	Markovitz WL
Louvian		0.586	6.17e-5	7.0e-14
GiadaMarsili			0.00042	8.4e-16
Markovitz				8.2e-27

Table 4: P values of t-test for K = 20 stocks.

As we can see in figure 19, for K = 25 stocks, Markowitz's free weights method tend to have a lower mean log return then other techniques. Markowitz's WL method has the highest mean log return. Louvian and Giadamarsili leads to comparable mean log returns and their difference is not statistically significant. On the other hand, both clustering techniques haves a higher mean log return if compared to Markowitz's methods and Markowitz's WL method dominance is statistically significant (see table 4). In table 1 we can observe sharpe rateos: clustering technique are able to outperform Markowitz's free weights method method but are beaten by Markowitz's WL method using this simple strategy.

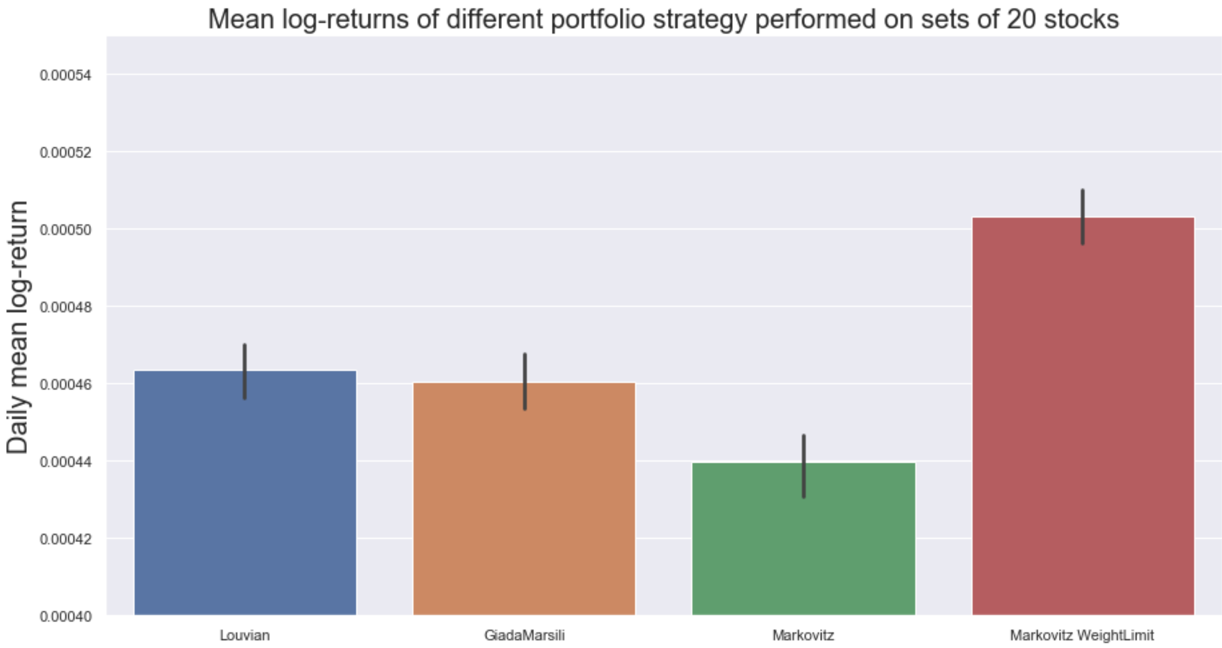


Figure 19: Mean log returns with different portfolio strategies on sets of 20 stocks. Markowitz's free weights optimization leads to lower mean daily retrurns. Markowitz's WL method has a higher mean return. The errorbars shows 95% confidence intervals computed with a bootstrapping technique.

We also noticed that given the high variance of daily log returns, in order to obtain statistically significant results, we had to repeat the simulations a large number of times. This limited our study because of the long running time of our simulations (multiple days). This was a concrete proof of the advantage that Louvian clustering has over MLE method because of the lower computational complexity.
These results demonstrate how important it is to have either a good method of forecasting expected returns or a strategy with low sensitivity. It is important to understand, that this is just one example of a very simple strategy that should not be seen as proof that clustering methods work best overall. If you are able to predict expected returns very accurately, then Markowitz's optimization performs better. Nevertheless, these results show that in some situations these methods can be better and therefore should be considered.

4 Outlook

In this project we have analyzed two clustering techniques for portfolio optimization. We have shown how clustering techniques, thanks to his highest stability with respect to variation in expected returns, can outperform optimization strategies in very simple cases. In our project we have tried to maintain flexibility for future use of our code. In fact, to keep the understanding simple, we have only tested with equalweights strategy, but it is possible to combine various strategies and to study their performance. It is possible to test the performance of a strategy by specifying two different techniques: an intra-cluster technique and an inter-cluster technique. One of the two will serve to determine the weights of the various stocks within the cluster, while the other will determine the weights of the various clusters. It's possible to chose between the following strategies: equal weights, risk parity, Markowitz's and it's variations. It is also possible to combine clustering strategies with Markowitz's ones. This will allow us, in the future, to test numerous different strategies and draw the most general conclusions possible. We are satisfied with the results obtained which still leave us intrigued to deepen this research as they do not show a method that predominates over the others in an absolute way. For future analysis it will be essential to include some more sophisticated expected return forecasting methods, to see if the advantage of having a low sensitivity impacts the results in such an essential way even if you have a better precision in forecasting expected returns. Finally, we are very pleased and proud to have been able to develop a complex project, thanks to the tools learned during this course.

5 Appendix

5.1 MLE Clustering

5.1.1 Industry

Asset No.		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1	Index	Homebuilding	Mining	Technology	Mining	Oil	Oil	Food	Food	Oil
2	Oil	Automobiles	Consumer Goods	Chemicals	Banking	Hotels	Banking	Oil	Homebuilding	Chemicals
3	Hotels	Index	Pharmaceuticals	Oil	Banking	Oil	Pharmaceuticals	Oil	Food	Oil
4	Hotels	Consumer Goods	Mining	Consumer goods	Aerospace	Consumer Goods	Oil	Oil	Aerospace	Real Estate
5	Aerospace	Mining	Food	Consumer Goods	Oil	Pharmaceuticals	Oil	Machinery	Mining	
6	Leisure	Hotels	Media	Banking	Technology	Media	Banking	Banking		
7	Real Estate	Banking	Technology	Software	Consumer Goods	Personal Services	Banking	Construction Materials		
8	Containers & Packaging	Office_Equipment	Food	Pharmaceuticals	Automobiles	Real Estate				
9	Banking	Oil	Real Estate	Oil	Consumer Goods	Machinery				
10	Real Estate	Real Estate	Consumer goods	Food	Oil	Oil				
11	Real Estate	Consumer Goods	Technology	Homebuilding	Pharmaceuticals	Office_Equipment				
12	Hotels	Aerospace	Pharmaceuticals	Chemicals	Food	Oil				
13	Consumer Goods	Automobiles	Automobiles	Banking	Textiles					
14	Mining	Multiline_Utilities	Machinery	Automobiles						
15	Water	Food	Consumer Goods							
16	Index	Oil	Hotels							
17	Technology	Banking	Banking							
18	Real Estate	Real Estate	Aerospace							
19	Chemicals	Real Estate	Insurance							
20	Machinery	Oil	Banking							
21	Oil	Water	Insurance							
22	Technology	Oil	Banking							
23	Technology	Banking	Consumer Goods							
24	Technology	Freight	Consumer goods							
25	Construction	Telecommunications	Real Estate							
26	Hotels	Consumer goods	Banking							
27	Banking	Banking								
28	Technology	Oil								
29	Commercial Services	Mining								
30	Media	Index								
31	Software									
32	Pharmaceuticals									
33	Banking									
34	Consumer Goods									
35	Paper									
36	Beverages									
37	Consumer Goods									
38	Aerospace									
39	Freight									
40	Consumer Goods									
41	Homebuilding									
42	Insurance									

Figure 20: MLE Industry clustering

5.1.2 Stock Exchange

Asset No.	Cluster 1		Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1	XNYS	XNYS	XNYS	XNAS	XNYS	XAMS	XNAS	XNYS	XNYS	XNYS
2	XLON	XNYS	XNYS	XNYS	XFRA	XNYS	XNYS	XNYS	XNYS	XAMS
3	XASX	XNYS	XCNQ	XLON	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS
4	XASX	XNYS	XWAR	XWAR	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS
5	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS	XNZE	
6	XNYS	XNYS	XNYS	XNYS	XLON	XNYS	XASX	XNYS		
7	XNYS	XNYS	XNYS	BMEX	XNYS	XNYS	XNYS	XNYS		
8	XNYS	XNYS	XASX	XNYS	XSWX	XFRA				
9	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS				
10	XNYS	XNYS	XNYS	XNYS	XNYS	XNYS				
11	XWAR	XNYS	XNYS	XNYS	XNYS	XNAS				
12	XNYS	XNYS	XTSX	XNYS	XNYS	XNYS				
13	XNZE	XPAR	XTOK	XNYS	XNYS					
14	XNYS	XNYS	XETR	XNYS						
15	XNYS	BMEX	XNYS							
16	XNYS	XNYS	XNYS							
17	XNYS	XNYS	XNAS							
18	XLON	XNYS	XNYS							
19	XNYS	XASX	XNYS							
20	XNYS	XNYS	XNYS							
21	XLON	XNYS	XNYS							
22	XNYS	XNYS	XLON							
23	XNYS	XNYS	XNYS							
24	XNAS	OTCM	XNYS							
25	XNYS	XLON	XNYS							
26	XLON	XNYS	XNYS							
27	XFRA	OTCM								
28	XNYS	XNYS								
29	XNYS	XNYS								
30	XNYS	XNYS								
31	XNYS									
32	XNYS									
33	XNYS									
34	XNYS									
35	XWAR									
36	XNYS									
37	XNYS									
38	XNYS									
39	XNYS									
40	XNYS									
41	XNYS									
42	XNYS									

Figure 21: MLE stock exchange

5.1.3 Market Capitalization

Asset No.	Cluster 1		Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1		4448500000	11147610000	1.7378E+11	1.5144E+11	809930898	9768444000	3.9856E+10	2.1975E+10	2.2621E+11
2	43976412	30782050000	2749939000	6.745E+10	9549803940	5922403000	1.2997E+11	103362000	3912875000	3.9453E+10
3	647603100	165643100	853079	8.7666E+10	4015817108	1672100000	1.3839E+11	1128723000	3.3657E+10	1.001E+11
4	29175490000	5919637000	222000000	4858127500	5441370000	1.0236E+11	9.5202E+10	4.6116E+10	9.8017E+10	4668224000
5	1.18316E+11	346630600	14509260000	2.1422E+10	2.9822E+10	1.9359E+11	2566868000	1502175000	223963153	
6	7765813000	2345124000	2.81537E+11	787841300	6.665E+10	6.9209E+10	301406200	8.3289E+10		
7	2296514000	714504840.2	30181840000	36586601.4	1.0374E+11	1.3663E+11	1.9763E+10	2.7547E+10		
8	31181290000	1167308000	14622750	3.3144E+11	5656793000	10640861				
9	1.1983E+11	551848000	10604260000	3581047000	4.3337E+11	6528455000				
10	623644300	81804240	969548500	4.0269E+10	26063000	2.5905E+11				
11	7970100	3.95855E+11	5724022000	8712860000	2.6423E+11	4041513000				
12	22545460000	1390000000	292590329	812966400	8.7055E+10	2982890000				
13	349555800	3834061680	6091177664	5058471000	958508700					
14	1406581000	22566520000	369260934	461476300						
15	3780411000	14480359.8	1.7415E+11							
16	169456398	1809193000	2.00314E+11							
17	709115100	27776180000	3004550000							
18	51270853.5	5777627000	270048400							
19	41852430000	16401600000	5071964000							
20	26147610000	315075000	7379159000							
21	1697795808	2183280000	2144000000							
22	80704120000	42008850000	12281822100							
23	25928740000	397341210.5	1.10888E+11							
24	56507640000	217404090	734599100							
25	6770583000	29054000000	8023766000							
26	158356755	5762443000	1.91307E+11							
27	81500000	106758000								
28	12656000000	9605368000								
29	1323493000	2762799000								
30	14746130000	426457293								
31	27494150000									
32	4.50358E+11									
33	4.67966E+11									
34	3594835000									
35	28082975									
36	2863559000									
37	5489806000									
38	98017260000									
39	25357850000									
40	1594222000									
41	17139820000									
42	2937974000									

Figure 22: MLE Market Cap

5.2 Stocks Used

TICKER	EXCHANGE_ABBREVIATION	MARKET_CAP_USD	OFFICIAL_NAME	INDUSTRY
^GSPC	XNYS		S & P 500	Index
GE	XNYS	1.03741E+11	GENERAL ELECTRIC COMPANY	Consumer Goods
IBM	XNYS	1.19866E+11	INTERNATIONAL BUSINESS MACHINES CORPORATION	Software
DIS	XNYS	2.81537E+11	THE WALT DISNEY COMPANY	Media
BA	XNYS	1.18316E+11	THE BOEING COMPANY	Aerospace
CAT	XNYS	1.11834E+11	CATERPILLAR INC.	Machinery
AA	XNYS	11147610000	ALCOA CORPORATION	Mining
DD	XNYS	41852430000	DuPont de Nemours, Inc.	Chemicals
XOM	XNYS	2.59052E+11	EXXON MOBIL CORPORATION	Oil
PG	XNYS	3.95855E+11	THE PROCTER & GAMBLE COMPANY	Consumer Goods
JNJ	XNYS	4.50358E+11	JOHNSON & JOHNSON	Pharmaceuticals
CVX	XNYS	2.26214E+11	CHEVRON CORPORATION	Oil
MCD	XNYS	2.00314E+11	MCDONALD'S CORPORATION	Hotels
MRK	XNYS	1.93588E+11	MERCK & CO., INC.	Pharmaceuticals
UTX	XSWX	1.17159E+11	RAYTHEON TECHNOLOGIES CORPORATION	Aerospace
MMM	XNYS	1.0236E+11	3M COMPANY	Consumer Goods
MO	XNYS	87054900000	Altria Group, Inc.	Food
HON	XNAS	1.43955E+11	HONEYWELL INTERNATIONAL INC.	Consumer Goods
ED	XNYS	30181840000	CONSOLIDATED EDISON, INC.	Technology
GT	XSWX	5656793000	THE GOODYEAR TIRE & RUBBER COMPANY	Automobiles
BMY	XNYS	1.38395E+11	BRISTOL-MYERS SQUIBB COMPANY	Pharmaceuticals
BP	XLON	87665881500	BP P.L.C.	Oil
LMT	XNYS	98017260000	LOCKHEED MARTIN CORPORATION	Aerospace
C	XNYS	1.1983E+11	CITIGROUP INC.	Banking
KR	XNYS	33657080000	THE KROGER CO.	Food
AET	XLON	43976412	AFENTRA PLC	Oil
XRX	XNAS	4041513000	XEROX HOLDINGS CORPORATION	Office_Equipment
F	XSWX	47710620000	FORD MOTOR COMPANY	Automobiles
AXP	XNYS	1.29969E+11	AMERICAN EXPRESS COMPANY	Banking
FRM	XASX	14622750	Farm Pride Foods	Food
EXC	XNAS	56507640000	EXELON CORPORATION	Technology
UNP	XNYS	1.6196E+11	UNION PACIFIC CORPORATION	Freight
EIX	XNYS	25928740000	EDISON INTERNATIONAL	Technology
LUV	XNYS	25357850000	SOUTHWEST AIRLINES CO.	Freight
CSX	XNAS	83396180000	CSX Corporation	Freight
SLB	XNYS	42008850000	Schlumberger N.V.	Oil
APA	XNAS	9768444000	APA CORPORATION	Oil
HUM	XNYS	59621920000	HUMANA INC.	Healthcare
DBD	XNYS	709115100	DIEBOLD NIXDORF, INCORPORATED	Technology
BC	XNYS	7765813000	BRUNSWICK CORPORATION	Leisure
PFE	XNYS	3.3144E+11	PFIZER INC.	Pharmaceuticals
COP	XNYS	95201590000	CONOCOPHILLIPS	Oil
LLY	XNYS	2.6423E+11	ELI LILLY AND COMPANY	Pharmaceuticals
PBI	XNYS	1167308000	PITNEY BOWES INC.	Office_Equipment
RSH	XASX	46261810	RESPIRI LIMITED	Healthcare
CLX	XNYS	21422350000	THE CLOROX COMPANY	Consumer Goods
AMD	XNAS	1.73775E+11	ADVANCED MICRO DEVICES, INC.	Technology
ABT	XNYS	2.48869E+11	ABBOTT LABORATORIES	Healthcare
TGT	XNYS	1.10888E+11	TARGET CORPORATION	Consumer Goods
DUK	XNYS	80704120000	DUKE ENERGY CORPORATION	Technology
APD	XNYS	67449680000	AIR PRODUCTS AND CHEMICALS, INC.	Chemicals
ASH	XNYS	6126756000	ASHLAND GLOBAL HOLDINGS INC.	Chemicals
MAS	XNYS	17139820000	MASCO CORPORATION	Homebuilding
WHR	XNYS	14714000000	WHIRLPOOL CORPORATION	Consumer Goods
NUE	XNYS	32623960000	NUCOR CORPORATION	Mining
VZ	XNYS	2.15123E+11	VERIZON COMMUNICATIONS INC.	Telecommunications
JPM	XNYS	4.67966E+11	JPMORGAN CHASE & CO.	Banking
T	XNYS	1.75669E+11	AT&T INC.	Telecommunications
ABM	XNYS	2749939000	ABM INDUSTRIES INCORPORATED	Consumer Goods
AP	XNYS	95468910	Ampco Pittsburgh	Mining
HD	XNYS	4.3337E+11	THE HOME DEPOT, INC.	Consumer Goods
BLL	XNYS	31181290000	BALL CORPORATION	Containers & Packaging
TE	XLON	29054000000	Telecom Egypt Company S.A.E.	Telecommunications
WFC	XNYS	1.91307E+11	WELLS FARGO & COMPANY	Banking
BT	XNAS	19035620	Bt Brands	Hotels
K	XNYS	21975130000	KELLOGG COMPANY	Food
CMI	XNYS	31200970000	CUMMINS INC.	Automobiles
PPL	XNYS	22566520000	PPL CORPORATION	Multiline_Utilities
LOW	XNYS	1.7415E+11	LOWE'S COMPANIES, INC.	Consumer Goods
WAG	XFRA	10640861	Swiss Estates AG	Real Estate
HSY	XNYS	39855540000	THE HERSHEY COMPANY	Food
DOV	XNYS	26147610000	DOVER CORPORATION	Machinery
MDP	XWAR	13075000	Medcamp	Real Estate
GCO	XNYS	969548500	GENESCO INC.	Consumer goods
FUR	XAMS	809930898	Fugro NV	Oil
CBU	XNYS	4015817108	COMMUNITY BANK SYSTEM, INC.	Banking
VHI	XNYS	812966400	VALHI, INC.	Chemicals
JWN	XNYS	3594835000	NORDSTROM, INC.	Consumer Goods

BDN	XNYS	2296514000	BRANDYWINE REALTY TRUST	Real Estate
BCS	XNYS	42268230000	BARCLAYS PLC	Banking
IDA	XNYS	5724022000	IDACORP, INC.	Technology
HRB	XNYS	4194978000	H & R BLOCK, INC.	Consumer Goods
CCE	XWAR	7970100	Clean&Carbon Energy SA	Real Estate
POM	XPAR	3834061680	Compagnie Plastic Omnium SE	Automobiles
DNP	XWAR	9002522500	DINO POLSKA S.A.	Food
TER	XNAS	26656100000	TERADYNE, INC.	Technology
BHP	XNYS	1.5144E+11	BHP GROUP LIMITED	Mining
DCI	XLON	51270853.5	Dolphin Capital Investors Limited	Real Estate
SYY	XNYS	40269160000	SYSCO CORPORATION	Food
GPS	XNYS	6590567000	THE GAP, INC.	Consumer Goods
ACG	XCNQ	853079	ALLIANCE GROWERS CORP.	Pharmaceuticals
GTY	XNYS	1455058000	GETTY REALTY CORP.	Real Estate
LEO	XETR	369260934	LEONI AG	Machinery
VGR	XNYS	1767454000	VECTOR GROUP LTD.	Food
SJT	XNYS	315075000	San Juan Basin Royalty Trust	Oil
USB	XNYS	83288740000	U.S. BANCORP	Banking
FRT	XNYS	10604260000	FEDERAL REALTY INVESTMENT TRUST	Real Estate
LEG	XNYS	5489806000	LEGGETT & PLATT, INCORPORATED	Consumer Goods
LM	XNYS	98017260000	LOCKHEED MARTIN CORPORATION	Aerospace
HSC	XNYS	1323493000	HARSCO CORPORATION	Commercial Services
IPG	XNYS	14746130000	THE INTERPUBLIC GROUP OF COMPANIES, INC.	Media
CW	XNYS	5441370000	CURTISS-WRIGHT CORPORATION	Aerospace
LG	XKRX	10850977200	LG Corp	Technology
MDC	XNYS	4448500000	M.D.C. HOLDINGS, INC.	Homebuilding
KBH	XNYS	3912875000	KB HOME	Homebuilding
CMC	XNYS	4412698000	COMMERCIAL METALS COMPANY	Mining
CRS	XNYS	1406581000	CARPENTER TECHNOLOGY CORPORATION	Mining
LEN	XNYS	35175680000	LENNAR CORPORATION	Homebuilding
CMO	XNZE	349555800	CMC	Consumer Goods
CTS	XNYS	1183286000	CTS CORPORATION	Technology
MYL	XASX	133077700	MALLEE RESOURCES LIMITED	Mining
MMC	XNYS	87760910000	MARSH & MCLENNAN COMPANIES, INC.	Insurance
SCG	XASX	16401600000	SCENTRE GROUP	Real Estate
TOL	XNYS	8712860000	TOLL BROTHERS, INC.	Homebuilding
ROL	XNYS	16832990000	ROLLINS, INC.	Consumer Goods
RLI	XNYS	5071964000	RLI Corp.	Insurance
TR	XNYS	2302241000	TOOTSIE ROLL INDUSTRIES, INC.	Food
SF	XNYS	7379159000	STIFEL FINANCIAL CORP.	Banking
PBT	XNYS	551848000	Permian Basin Royalty Trust	Oil
THO	XNYS	5762443000	THOR INDUSTRIES, INC.	Consumer goods
RVT	XNYS	2103872295	Royce Value Trust	Banking
PKE	XNYS	270048400	PARK AEROSPACE CORP.	Aerospace
SOR	XNYS	397341210.5	Source Capital	Banking
NRT	XNYS	103362000	North European Oil Royalty Trust	Oil
OXM	XNYS	1715191000	OXFORD INDUSTRIES INC	Textiles
TAI	OTCM	217404090	Thai Airways International Public Company Limited	Freight
MSB	XNYS	346630600	Mesabi Tr	Mining
CAH	XNYS	14509260000	CARDINAL HEALTH, INC.	Food
VNO	XNYS	8023766000	VORNADO REALTY TRUST	Real Estate
VMC	XNYS	27546910000	VULCAN MATERIALS COMPANY	Construction Materials
TEG	XFRA	4110081900	TAG Immobilien AG	Real Estate
UL	XNYS	1.3663E+11	UNILEVER PLC	Personal Services
UGI	XNYS	9605368000	UGI CORPORATION	Oil
WRI	XTSX	1250440	Waseco Resources Inc.	Mining
UNF	XNYS	3970020000	UNIFIRST CORPORATION	Commercial Services
LZB	XNYS	1594222000	LA-Z-BOY INCORPORATED	Consumer Goods
TFX	XNYS	15387630000	TELEFLEX INCORPORATED	Healthcare
FRX	XASX	24329120	FLEXIROAM	Telecommunications
USG	XNZE	223963153	American Pacific Mining Corp	Mining
CCL	XNYS	22545460000	Carnival Corp	Hotels
MU	XNYS	2345124000	MANCHESTER UNITED PLC	Hotels
ZTR	XNYS	426457293	Virtus Total Return Fund	Index
PMM	XNYS	405694486.5	Putnam Managed Municipal Income Trust	Banking
CXE	XNYS	169456398	MFS High Income Municipal Trust	Index
EOG	XNYS	51973540000	EOG RESOURCES, INC.	Oil
STJ	XLON	12281822100	ST. JAMES'S PLACE PLC	Banking
DSM	XAMS	39452698200	Royal DSM N.V.	Chemicals
RF	XNYS	20781560000	REGIONS FINANCIAL CORPORATION	Banking
MKC	XNYS	25814120000	MCCORMICK & COMPANY, INCORPORATED	Food
MGA	XNYS	30782050000	Magna International Inc.	Automobiles
IGT	XNYS	5922403000	INTERNATIONAL GAME TECHNOLOGY PLC	Hotels
MCY	XNYS	2937974000	MERCURY GENERAL CORPORATION	Insurance
OII	XNYS	1128723000	OCEANEERING INTERNATIONAL, INC.	Oil
VMI	XNYS	5316658000	VALMONT INDUSTRIES, INC.	Construction
ONB	XNAS	3004550000	OLD NATIONAL BANCORP	Banking
PVA	BMEX	14480359.8	Pescanova SA	Food
CWT	XNYS	3780411000	CALIFORNIA WATER SERVICE GROUP	Water
MSA	XNYS	5919637000	MSA Safety Incorporated	Consumer Goods

WOR	XNYS	2762799000	WORTHINGTON INDUSTRIES, INC.	Mining
FUL	XLON	158356755	Fulham Shore	Hotels
WTS	XNYS	6528455000	WATTS WATER TECHNOLOGIES, INC.	Machinery
BRS	XWAR	222000000	Boryszew S.A.	Mining
CPF	XNYS	787841300	CENTRAL PACIFIC FINANCIAL CORP.	Banking
CPK	XNYS	2566868000	CHESAPEAKE UTILITIES CORPORATION	Oil
SSP	XNAS	1596846000	THE E.W. SCRIPPS COMPANY	Media
SR	XNYS	3374609000	Spire Inc.	Oil
KWR	XNYS	4129688000	QUAKER CHEMICAL CORPORATION	Chemicals
MOD	XNYS	523202800	MODINE MANUFACTURING COMPANY	Automobiles
ENB	XNYS	1.00103E+11	Enbridge Inc.	Oil
EV	XTSX	13641690	ERIN VENTURES INC.	Mining
STC	XNYS	2144000000	STEWART INFORMATION SERVICES CORPORATION	Insurance
INT	XNYS	1672100000	WORLD FUEL SERVICES CORPORATION	Oil
MLI	XNYS	3405288000	MUELLER INDUSTRIES, INC.	Machinery
NBR	XNYS	668270300	NABORS INDUSTRIES LTD.	Oil
IO	XNYS	26063000	ION GEOPHYSICAL CORPORATION	Oil
GFI	XNYS	9709025000	Gold Fields Limited	Mining
AES	XNYS	16201150000	THE AES CORPORATION	Technology
XL	XNYS	461476300	XL FLEET CORP	Automobiles
MTB	XNYS	19763380000	M&T BANK CORPORATION	Banking
NTX	BMEX	36586601.4	Netex	Software
TOT	XTSE	265320000	TOTAL ENERGY SERVICES INC.	Oil
EEP	BMEX	6390486.6	EuroEspes	Healthcare
ROP	XNYS	51883800000	ROPER TECHNOLOGIES, INC.	Software
GVA	XNYS	1773495000	GRANITE CONSTRUCTION INCORPORATED	Construction
CRT	XNYS	75900000	Cross Timbers Royalty Trust	Index
BRO	XNYS	19848960000	BROWN & BROWN, INC.	Insurance
WBS	XNYS	5058471000	WEBSTER FINANCIAL CORPORATION	Banking
TNC	XNYS	1502175000	Tennant Company	Machinery
MYF	XNYS	714504840.2	First Western Financial	Banking
DVN	XNYS	29821850000	DEVON ENERGY CORPORATION	Oil
CVC	XASX	301406200	CVC	Banking
EGP	XNYS	9269911000	EASTGROUP PROPERTIES, INC.	Real Estate
SJW	XNYS	2183280000	SJW GROUP	Water
SCL	XNYS	2786059000	STEPAN COMPANY	Chemicals
PEI	XNYS	81804240	PENNSYLVANIA REAL ESTATE INVESTMENT TRUST	Real Estate
UTL	XNYS	734599100	Unitil	Consumer goods
CBK	XFRA	9549803940	Commerzbank Aktiengesellschaft	Banking
MYJ	XNYS	386044850.8	BlackRock MuniYield New Jersey Fund	Index
DHI	XNYS	38665630000	D.R. HORTON, INC.	Homebuilding
FCF	XNYS	1527457000	FIRST COMMONWEALTH FINANCIAL CORPORATION	Banking
KMP	XWAR	28082975	Przedsiebiorstwo Produkcyjno Handlowe Kompap SA	Paper
SM	XNYS	3581047000	SM ENERGY COMPANY	Oil
MNP	XNYS	165643100	Western Asset Municipal Partners Fund Inc.	Index
CDR	XWAR	4858127500	CD PROJEKT S.A.	Consumer goods
MS	XNYS	1.7614E+11	MORGAN STANLEY	Banking
ALG	XASX	647603100	ARDENT LEISURE GROUP LIMITED	Hotels
HR	XNYS	4668224000	HEALTHCARE REALTY TRUST INCORPORATED	Real Estate
RIG	XNYS	1809193000	Transocean Ltd.	Oil
ALL	XASX	29175490000	ARISTOCRAT LEISURE LIMITED	Hotels
YPF	XNYS	2982890000	YPF SA	Oil
KOF	XNYS	2863559000	Coca-Cola Femsa SAB de CV	Beverages
MOV	XNYS	958508700	MOVADO GROUP, INC.	Textiles
IT	XNYS	27494150000	GARTNER, INC.	Software
CBL	XNYS	623644300	CBL & ASSOCIATES PROPERTIES, INC.	Real Estate
SU	XNYS	46115780000	SUNCOR ENERGY INC.	Oil
SUI	XNYS	24349580000	SUN COMMUNITIES, INC.	Real Estate
MAA	XNYS	26417340000	MID-AMERICA APARTMENT COMMUNITIES, INC.	Real Estate
RYN	XNYS	5777627000	RAYONIER INC.	Real Estate
RKT	XNYS	27776180000	Rocket Companies	Banking
EXP	XNYS	6770583000	EAGLE MATERIALS INC.	Construction
GXP	XFRA	81500000	GXP GERM.PROP.AG	Banking
PL	XNYS	1390000000	Planet Labs	Aerospace
USA	XNYS	1862000000	Liberty All-Star Equity Fund	Index
WGL	XNYS	4922000	Western Gold Exploration Limited	Mining
BSP	XNYS	669819000	Franklin BSP Realty Trust, Inc.	Index
HCP	XNYS	12656000000	HashiCorp	Technology
KYO	XTOK	22331434021	Kyocera Corporation	Technology
KWK	XTOK	6091177664	Kawasaki	Automobiles
BXS	XNYS	5060000000	Blackstone Secured Lending Fund	Index
TWX	XNYS	69209000000	Warner Bros	Media
DRL	XLON	1697795808	The Drilling Company	Oil
ISH	XTSX	292590329	Inner Spirit Holdings Ltd	Pharmaceuticals
ELX	XLON	66650000000	Electrolux	Technology
TYC	OTCM	106758000	Calvin B. Taylor Bankshares	Banking

References

[1] H. E., *Quantitative Expression of Cultural Relationships.* University of California Publications in American Archaeology and Ethnology, 1932.

[2] R. Mantegna, *Hierarchical structure in financial markets.* The European Physical Jorunal, 1999.

[3] G. Martia, F. Nielsenc, M. Bińkowskid, and P. Donnat, *A review of two decades of correlations, hierarchies, networks and clustering in financial markets.* Progress in Information Geometry, 2020.

[4] V. B. et al., *Fast unfolding of communities in large networks.* Journal of Statistical Mechanics: Theory and Experiment, 2008.

[5] M. MacMahon and D. Garlaschelli, *Community detection for correlation matrices.* Physical review X, 2015.

[6] L. Giada and M. Marsili, *Algorithms of maximum likelihood data clustering with applications.* Physica, 2002.

[7] Z. Ren, *Portfolio Creation using Clustering Methods.* Univesrity of Worcester, 2005.

[8] E. Qian, *Risk Parity Portfolios: Efficient Portfolios Through True Diversification.* Panagora, 2005.

[9] H. Markowitz, *Portfolio Selection.* The Journal of Finance, 1991.

[10] L. Bachelier, *Theory of Speculation.* 1900.

[11] F. Black and R. Litterman, *Black-Litterman Global Asset Allocation Model.* Goldman Sachs Asset Management, 1990.

[12] B. Chabot, E. Ghysels, and R. Jagannathan, *Momentum Trading, Return Chasing, and Predictable Crashes.* SSRN, 2014.