

A proposal of a methodological framework with experimental guidelines to investigate clustering stability on financial time series

Gautier Marti, Philippe Very
Philippe Donnat
Hellebore Capital Management
63 Avenue des Champs-Élysées
firstname.lastname@helleborecapital.com

Frank Nielsen
Laboratoire d'Informatique de l'École polytechnique
Palaiseau, FRANCE
nielsen@lix.polytechnique.fr

Abstract—We present in this paper an empirical framework motivated by the practitioner point of view on stability. The goal is to both assess clustering validity and yield market insights by providing through the data perturbations we propose a multi-view of the assets' clustering behaviour. The perturbation framework is illustrated on an extensive credit default swap time series database available online at www.datagrapple.com.

I. INTRODUCTION

Clustering, the task of grouping a set of objects in such a way that objects in the same group, also called cluster, are more similar to each other than those in different groups, is not yet a common tool for financial time series analysis. However, its application area seems wide: portfolio diversification [1] and risk analysis, investment strategies such as statistical arbitrage. This unsupervised machine learning technique suffers from several drawbacks which must be dealt with before being adopted by the financial community. These algorithms act as black boxes whose properties such as consistency, i.e. the probability that a given algorithm outputs the target clustering converges to 1, and rate of convergence are generally unknown in the asymptotic regimes which matter in finance, i.e. Random Matrix Theory Asymptotics [2], High Dimension Moderate Sample Size Asymptotics (HDMSS) and High Dimension Low Sample Size Asymptotics (HDLSS) [3]. Besides lacking from these theoretical properties for now, clusterings obtained are much dissimilar from one algorithm to another when applied on financial time series [4], [5]. Even worse, adding small noise to a given sample and applying the same algorithm on the original one and its perturbed version yields different clusters. These current shortcomings prevent robust cluster-based applications in finance where data is noisy observations of the underlying phenomenon. For instance, a trading strategy based on clustering events such as emergence, combination, split and decay [6], [7] of clusters is flawed if these events spuriously occur due to the algorithm unstability.

Concerning the object of study, namely Credit Default Swaps (CDS), they are vanilla credit derivatives (cf. [8] for an introduction to derivatives and [9] for a thorough treatment of credit derivatives and CDS mathematical modelling) whose purpose is to transfer credit risk from a counterparty to another. They may be considered as an insurance against the default (or other credit events defined precisely by an ISDA contract)

of the underlying entity, usually a corporate, a financial or a sovereign entity. The price of this insurance can thus be viewed as a fear gauge expressing the market apprehension of an entity default risk: higher the risk, higher the price to pay to get insured. The CDS market (being an over-the-counter (OTC) market) was accused to lack transparency, today it is still difficult to gather public information on this market. We can mention DTCC which reports traded volumes once a week, and DataGrapple which displays historical price time series of 700 entities from 2006 to present updated on a daily basis.

A. Related Work

Clustering financial time series was mainly explored in econophysics [1], [10], [11] where it is considered as a competitive way to the Random Matrix Theory (RMT) [2], [12] to filter the correlation matrix. It was noticed (for instance, in [5]) that clustering algorithms recover roughly industrial classification benchmarks (ICB). Authors propound to benchmark these algorithms against an ICB, and measure the similarity between a resulting clustering and the chosen ICB using the Adjusted Rand Index (ARI). They consider this measure as a proxy of the information filtered by the methods, and thus consider it as a way to perform model selection by opting for the clustering method with the highest score. Yet, a statistical analysis such as clustering can produce clusters which may not exist in an industrial classification: for example, the PIIGS cluster (assets from Portugal, Ireland, Italy, Greece and Spain) during the European debt crisis. In [5], authors also use bootstrapping [13] to assess the robustness of their clustering algorithm using the intuitive but moot idea of clustering stability [14], i.e. the reproducibility of the clustering when data is slightly perturbed. Clustering stability for model selection or validity assessment [15] is indeed a hot topic in the machine learning literature: [16] warn against its irrelevant use as stability only depends on the uniqueness of the clustering objective function minimizer for large sample size, yet [17] advocate that this criterion remains useful in the case of finite possibly small samples. From the quantitative analyst or trader perspective, dynamical stability of models, i.e. stability to online perturbations arising from streaming data, is required for being confident using them. A clustering should only change when a meaningful event happens in the market. Despite being an important notion for practitioners, few works

have dealt with the dynamical stability properties of clusters computed on financial time series [18]. Clustering CDS has not been widely explored in the literature since data is scarce and difficult to obtain. [19] investigate the differences in the main data sources employed by researchers and policymakers and conclude that credit default swap databases are not all equal. Some of these databases are built using hypothetical prices resulting from an average of data collected but on which, possibly, no trade could have been done. We will use the database from DataGrapple which is built tick by tick from many market dealers. This allows to make a synthetic order book of best bid / best offer across the dealers' prices at a given time t improving the confidence that a trade could have effectively been done at this level at time t .

B. Contributions

We propose an empirical framework to investigate clustering on financial time series and obtain insights from the perturbations which are motivated by financial applications. To illustrate the proposed perturbations, we apply them to study a large CDS dataset of market prices. We compare on this dataset the clusterings obtained using several distance matrices D_{ij} :

- $D_{ij}^P = (1 - \rho_{ij})/2$, where ρ_{ij} is the Pearson correlation coefficient,
- $D_{ij}^S = (1 - \rho_{S_{ij}})/2$, where $\rho_{S_{ij}}$ is the Spearman correlation,
- D_{ij}^E is an Euclidean distance,
- D_{ij}^G is a distance designed for working on i.i.d. random processes [20].

Our experimental study advocates that:

- if dependence between assets is the only focus of study, D_{ij}^S (Spearman) should be preferred to D_{ij}^P (Pearson); D_{ij}^E should not be used for comparing financial time series; finally, D_{ij}^G is the most versatile distance and behaves well under all proposed perturbations, it mostly behaves like D_{ij}^S and should be preferred;
- clustering can be persistent through a wide range of perturbations provided that an appropriate distance is leveraged by the clustering algorithm.

Besides, we observe some interesting facts about the CDS market and illustrate clustering stability by leveraging the Sankey diagram from the Data-Driven Documents library.

II. PERTURBATION FRAMEWORK FOR FINANCIAL TIME SERIES

Perturbations can be performed both on prices (some of the T values of each time series) or on assets (some of the N time series themselves). Concretely, these perturbations consist in modifying row-wise or column-wise the $N \times T$ data matrix X .

A. Time Perturbation

We provide below a list of perturbations concerning some of the T time series values. We explain their motivations arising from financial concerns and what we can learn from them by analyzing the clustering stability.

Sliding Window: **Motivation:** Dynamical stability of models is a requirement for trading and risk information systems. For example, value at risk (VaR), an estimated amount of money so that the potential loss of a portfolio over a given timespan should not exceed, is computed on a moving window and updated with respect to the asset prices stream. With no trading in the portfolio, and in a stationary regime, VaR should not vary too much.

Definition: Given a window width W and a step size S , clustering is performed on $X_{:, [t_{\text{cur}}, t_{\text{cur}}+W[}$ and $X_{:, [t_{\text{cur}}+S, t_{\text{cur}}+S+W[}$, then current time t_{cur} is updated $t_{\text{cur}} := t_{\text{cur}} + S$, and so on.

Insight: A clustering that strongly differs from one time to another when the market seems in a steady regime should be rejected since very sensitive to noise, i.e. small insignificant market variations. If confidence is high in the methodology, modification of clusters may be a signal that market structure is changing (for instance, end of a crisis and decrease in global correlation).

Odd vs. Even: **Motivation:** A clustering algorithm applied on two samples describing the same phenomenon should yield the same results. How to obtain two of these samples? The goal is to split the sample in two while mitigating the effect of non-stationarity, seasonality, end-of-the-week trading activity, meetings and announcements from the ECB or the FED generally happening on Friday.

Definition: We define $X^{(1)} = \{X_{:,t} \mid t \text{ is odd}\}$ and $X^{(2)} = \{X_{:,t} \mid t \text{ is even}\}$, i.e. we build the sample of the odd trading days and the sample of the even trading days. Since the trading week lasts 5 days, we alleviate the aforementioned statistical biases. Clustering is performed independently on $X^{(1)}$ and on $X^{(2)}$.

Insight: If not stable, the clustering method should be rejected.

Economic Regimes: **Motivation:** Since the economic context can change dramatically, financial time series do not evolve in a steady regime in the long run. It makes sense to split the timespan into different periods where the statistical regime can be considered stationary.

Definition: We partition the sample into M subsamples $X = \sqcup_{i=1}^M X^{(i)}$, where $X^{(i)} = X_{:, [t_i, t_{i+1}[$ and the $(t_i)_{i=1}^{M+1}$ delimit time intervals. The breakpoints $(t_i)_{i=1}^{M+1}$ can be chosen guided by market understanding or computed using a dynamical changes and regime detection algorithm.

Insight: We can study whether the clustering structure is persistent throughout different economic regimes, and how strongly it is. If not, which are the periods concerned and how steep is the change? Which assets are involved? Why do they switch from clusters? What happened to them and their clusters? However, we must keep in mind that it is difficult to separate the signal from the noise of the clustering methodology.

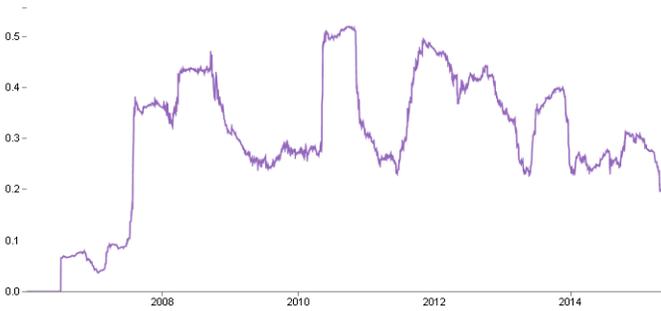


Fig. 1: Mean Correlation Dynamics computed on the whole CDS dataset from 2006 to 2015 using a 6-month sliding window

Heart vs. Tails: Motivation: Does the market under stress share a common clustering structure with the market during uneventful periods?

Definition: Let $\bar{X}_t = \frac{1}{N} \sum_{i=1}^N X_{i,t}$ be the mean time series of the market. let Q_1 be the lower quartile and Q_3 be the upper quartile. We define $\mathcal{T} = \{t \mid \bar{X}_t \leq Q_1 \vee \bar{X}_t \geq Q_3\}$ and $\mathcal{H} = \{t \mid \bar{X}_t \notin \mathcal{T}\}$ corresponding to times having market values in the tails and in the heart respectively. Then we split the sample X in the two following subsamples $X^{(1)} = \{X_{:,t} \mid t \in \mathcal{T}\}$ and $X^{(2)} = \{X_{:,t} \mid t \in \mathcal{H}\}$ on which we apply the same clustering algorithm.

Insight: Although it is difficult to anticipate changes of the market behaviour, in period of stress all assets tend to be simultaneously affected by macroeconomic tensions which usually induces a significant increase in correlation between them (cf. Fig. 1). Thus, correlation should be less discriminating and a correlation-based clustering might be unstable with respect to this perturbation.

Multiscale: Motivation: Markets prices can be monitored from a high frequency sampling (tick by tick or minute by minute) to much lower frequency (from hours to hours, days by days or on a weekly basis). The sampling frequency used is linked to the type of trading, from high-frequency trading (HFT) and algorithmic trading to long-term investments. Is the clustering structure persistent throughout a wide range of time scales or does it strongly depend on the sampling?

Definition: From X we can build M datasets $X^{(i)} = X_{:,s_i}$, $i = 1, \dots, M$ where s_i are regularly-spaced multiscale subsample of $\{1, \dots, T\}$.

Insight: Ideally, in the perspective of building a risk system, we would like that the choice of risk factors is independent of the time scale used for the analysis. This perturbation allows us to verify to which extent clustering is multiscale persistent.

Maturities: Motivation: Fixed-income assets such as bonds, swaps and CDS for instance, have a lifespan called their maturity. Several products with different maturities (say an insurance against the default of a corporate for 1, 2, 3, 5, 7 or 10 years) can concern the same entity. Since the underlying risk is the same, we would like similar clusterings.

Definition: We get from the market several time series dataset $X^{(i)}$, one $N \times T$ data matrix for each quoted maturity.

Insight: We can either reject a clustering method which yields to unstable clusters or investigate why a particular maturity has a different clustering structure compared to the others.

Term Structure: Motivation: The term structure is the set of all quoted maturities. Clustering term structure could lead to a more meaningful result than clustering separately each maturity.

Definition: For clustering term structures, one need a specific distance. To our knowledge, the problem of obtaining a proper one which captures the whole information (e.g. dynamics, distribution and correlation of its distortions) has not been addressed. Here, we give a simple one for clustering CDS term structures at a given date t . A CDS probability of default $P(t)$ can be viewed as a cumulative distribution function on \mathbf{R}^+ . Indeed, the probability of default is increasing, the probability of instantaneous default is 0, and at infinity all entities will eventually default. Thus, $f(t) = \partial P(t)/\partial t$ defines a probability density function on \mathbf{R}^+ , and since $\int_{\mathbf{R}^+} f(t) dt = 1$, $\sqrt{f(t)}$ is a unit vector in $L^2(\mathbf{R})$. The inner product between two unit vectors defines an angle ϕ which is the distance between two term structures. Given two term structures P_1, P_2 and f_1, f_2 such that $f_i(t) = \partial P_i(t)/\partial t$, their distance ϕ can be written $\cos \phi = \int_{\mathbf{R}^+} \sqrt{f_1(t)} \sqrt{f_2(t)} dt = 1 - H^2(\sqrt{f_1(t)}, \sqrt{f_2(t)})$, where H is the Hellinger distance.

Insight: For entities near default, the term structure should be inverted, i.e. the market anticipates a renormalisation if these entities survive. For entities having seemingly no troubles, the quoted term structure should mirror the debt term structure of these entities. Some industries have a particular debt structure (short term debt for financials, long term debt for basic materials and industrials). Part of this information should also be captured by correlation between assets on a given maturity.

B. Population Perturbation

To the presented time-based perturbations, we add the following two population-based perturbations on the set of assets: **increasing/decreasing the number of entities** and **adding entities with imputed historical prices**. These perturbations can be easily motivated: new companies emerge regularly and some others disappear from the market. The clustering structure should not radically change when adding or removing entities from the clustering perimeter. When new companies are created and introduced in the market, they have not much history. It may be necessary to impute missing data based, for instance, on a clustering methodology. We would like to verify that adding synthetic time series built from existing ones to the clustering perimeters does not change the original clustering structure. The clustering structure should be robust to the statistical engineering performed to impute missing data or clean their poor quality.

III. INSIGHTS FROM BENCHMARKING ON STABILITY

A. CDS Dataset

We apply the proposed perturbation framework on a dataset of Credit Default Swap (CDS) time series which can be seen at www.datagrappe.com. This dataset have 450 entities which

have complete historical daily prices from January 2006 to today, i.e. more than 2300 daily prices for a given asset (say Nokia Oyj) on a given maturity (say a 5 year contract), and 250 more with missing data (for instance, Numericable Group S.A.) whose missing historical prices were imputed using a machine learning algorithm. Since credit default swaps are traded over-the-counter, closing time for fixing prices can be arbitrarily chosen, here 5pm GMT, i.e. after the London Stock Exchange trading session. This synchronous fixing of CDS prices avoids spurious correlations arising from different closing times. For example, the use of close-to-close stock prices artificially overestimates intra-market correlation and underestimates inter-market dependence since they have different trading hours [21]. Moreover, besides being publicly available, another benefits of this dataset is that historical prices are mid-market data computed from a synthetic real-time order book of best bid / best offer that are proposed by the main CDS market makers instead of averaged consensus prices on which it could have been impossible to trade.

B. Preprocessing of Financial Time Series

Given N time series of prices $P_i(t)$, $i = 1, \dots, N$, $t = 1, \dots, T$, the standard approach consists in working with the time series of log-returns $\Delta \log P_i(t) = \log P_i(t+1) - \log P_i(t)$. This transform mitigates the risk of spurious correlation by empirically stationarizing the considered time series besides normalizing the variations. Thus, clustering the N time series $\Delta \log P_i(t)$, $t = 1, \dots, T$ using a distance based on Pearson correlation or an Euclidean distance may be more relevant. In [20], authors claim that the standard Euclidean distance should not be used for clustering financial time series. For instance, assuming that variations of two time series X and Y are independent and identically distributed according to $\mathcal{N}(\mu_X, \sigma_X^2)$ and $\mathcal{N}(\mu_Y, \sigma_Y^2)$, we have $\mathbf{E}[(X - Y)^2] = (\mu_X - \mu_Y)^2 + (\sigma_X - \sigma_Y)^2 + 2\sigma_X\sigma_Y(1 - \rho(X, Y))$. If X and Y are independent, then $\rho(X, Y) = 0$, and $\mathbf{E}[(X - Y)^2] = (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2$. Now, suppose that $\mu_X = \mu_Y$ and $\sigma_X = \sigma_Y$, the distance $2\sigma_X^2$ between X and Y grows proportionally to the variance which is not a desirable behaviour for two time series whose variations have the same distribution. Based on the observation that the Euclidean distance mixes correlation and distribution information but inappropriately for their purpose, authors have designed a distance of the form $D_{\text{correlation}} + D_{\text{distribution}}$ for comparing i.i.d. random processes both on correlation and distribution information. Using their distance we can work directly on $\Delta P_i(t) = P_i(t+1) - P_i(t)$ since the log-transform is useless: both the dependency (based on Spearman correlation) and distribution (Hellinger) distance employed are invariant to monotonic transforms of the data. Thus, we will work on $\Delta \log P_i(t)$ when clustering using D^P and D^E , and on $\Delta P_i(t)$ when clustering using D^S and D^G . In the remaining, we apply for both experiments a hierarchical clustering, namely weighted linkage, with a constant height cut to obtain $K = 16$ clusters; results presented still holds for various values of K and different clustering algorithms.

C. Comparison of Distances using the Perturbation Framework

In this section, we leverage the proposed perturbation framework to test four distances used for clustering financial

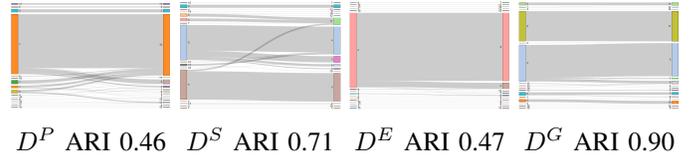


Fig. 2: Stability to Odd vs. Even perturbation and the associated ARI showing a better stability of D^G -based clustering; partitions obtained from D^S -based clustering are rather similar but less stable

time series. We also observe some stylized facts about the CDS market. The four distances D^P , D^S , D^E and D^G are essentially distances based on Pearson correlation, Spearman correlation, Euclidean distance and a recently introduced distance based on correlation and distribution respectively. We illustrate the clustering stability with a data visualization, a Sankey diagram, which highlights the dissimilarities between partitions [22].

In Fig. 2, we display the stability results on the Odd vs. Even experiments. For each distance, we have displayed the partitions obtained on the odd trading days sample (left) and on the even trading days sample (right). A grey link binds a given asset in the left partition to the same asset in the right partition. Thus, a perfectly stable clustering is displayed by a one-to-one correspondence between left and right clusters. Diverging edges highlight mismatches between partitions, hovering on the edges shows the assets switching from clusters. In this experiment, these can be assets with an unusual history, for instance they may have encountered a strong variation on a particular day due to a merger (M&A), a catastrophe or a fraud. But, of course, a cluster switch can happen due the clustering method shortcomings.

The Heart vs. Tails experiment displayed in Fig. 3 shows an interesting stylized fact about the CDS market. Clustering on correlation (D^S and D^P) is not stable at all with respect to this perturbation. This means that the sample of the strongest moves in the market has a totally different clustering structure than the sample of the mildest moves when considering only correlation. This can be explained since when the market is stressed, macroeconomic tensions tend to affect all the participants and correlation between assets becomes significantly higher and similar for all assets, thus becomes uninformative. This claim is supported by the fact that D^S based on the Spearman correlation (correlation between ranks) performs the worst, whereas D^P , based on the Pearson correlation measure known to be decreased by fat-tailed variations, achieves a better stability since this correlation-based distance discriminates unintentionally on distributions. For high values of ρ , which is the case in stressed period, D^E discriminates on the mean and variance of the variations, so performs better than the correlation-based distances. Finally, D^G which intentionally works on both information can leverage the distribution information and obtain a rather stable clustering between the stress periods and the more quiet ones.

In Fig. 4, we display results of the Maturity experiment. For each clustering, we show 5 partitions corresponding to clustering the 1,3,5,7,10-year CDS. We can notice that the partition corresponding to the 1-year CDS is the less stable

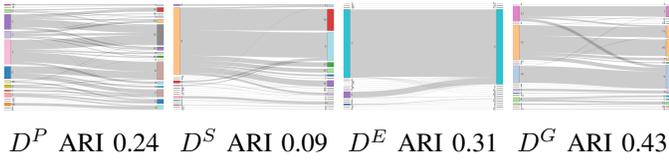


Fig. 3: Stability to Heart vs. Tails perturbations for different distances and the associated ARI

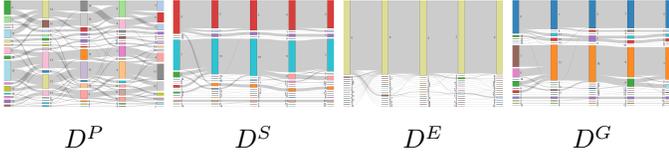


Fig. 4: Stability to Maturity perturbations

whatever the distance used. This can be explained by the relative illiquidity of the 1-year maturity compared to the others yielding to scarce and noisy quotes from the market makers. Stability is high for D^S and D^G and abnormally low for D^P and D^E while information is essentially the same.

The Fig. 5 depicts results of the Multiscale experiments. 6 partitions are displayed for each distance corresponding to the clusterings obtained by considering respectively 1,2,4,8,16,32 trading days variations. We can observe as a stylized fact that the clustering structure is persistent up to a weekly sampling, and that the clustering structure is essentially determined by correlations as advocated by the high stability achieved by D^S and D^P . D^G , once again, is relatively stable leveraging its correlation part which is similar to D^S .

We finally conclude this empirical study with the Economic Regimes perturbations. In Fig. 6, we display 4 partitions corresponding to the clusterings obtained on different economic periods. From left to right, the pre subprime crisis period 2006-2007, the subprime crisis period 2008-2009, the European debt crisis 2011-2012 and the quantitative easing 2013-2014. We can notice in Fig. 6 that the period 2006-2007 yields very different clusters compared to what follows. Indeed, looking at Fig. 1, we observe that correlation in the market was very low. Except clustering with D^G , clusters obtained with the other methods are not stable. The partitions and their stability scores obtained from the D^G -based clustering agree with previous remarks: pre-crisis period was much different, the clustering structure is the same during both crises, and now that correlation is decreasing and that quantitative easing is at work the clustering structure of the market is changing.

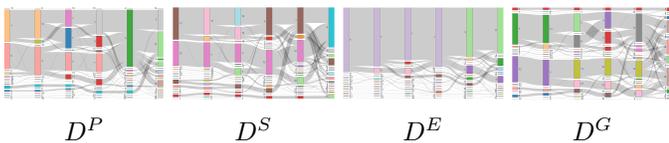


Fig. 5: Stability to Multiscale perturbations

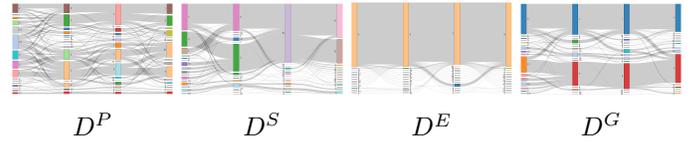


Fig. 6: Stability to Economic Regimes perturbations

IV. DISCUSSION

In this paper, we have suggested an empirical framework for both assessing clustering validity and investigating the clustering properties of financial time series. This empirical study allows to verify known stylized facts about the financial markets. Much research work needs to be done before clustering can become a technology in trading rooms or an accepted tool for economic applied research. With this experimental study, we hope to have aroused curiosity from the clustering community, and that further progress will be achieved in the field of clustering financial time series using a framework based on this one for systematically benchmarking the proposed algorithms taking into account the wishes of practitioners.

REFERENCES

- [1] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna, "Cluster analysis for portfolio optimization," *Journal of Economic Dynamics and Control*, vol. 32, no. 1, pp. 235–258, 2008.
- [2] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 3, no. 03, pp. 391–397, 2000.
- [3] P. Borysov, J. Hannig, and J. Marron, "Asymptotics of hierarchical clustering for growing dimension," *Journal of Multivariate Analysis*, vol. 124, pp. 465–479, 2014.
- [4] V. Lemieux, P. S. Rahmdel, R. Walker, B. Wong, and M. Flood, "Clustering techniques and their effect on portfolio formation and risk analysis," in *Proceedings of the International Workshop on Data Science for Macro-Modeling*. ACM, 2014, pp. 1–6.
- [5] N. Musmeci, T. Aste, and T. Di Matteo, "Relation between financial market structure and the real economy: Comparison between clustering methods," *Available at SSRN 2525291*, 2014.
- [6] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu, "Graphscope: parameter-free mining of large time-evolving graphs," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 687–696.
- [7] C. Tantipathananandh and T. Y. Berger-Wolf, "Finding communities in dynamic social networks," in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 1236–1241.
- [8] J. C. Hull, *Options, futures, and other derivatives*. Pearson Education, 2006.
- [9] D. O’Kane, *Modelling single-name and multi-name credit derivatives*. John Wiley & Sons, 2011, vol. 573.
- [10] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, "A tool for filtering information in complex systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 30, pp. 10 421–10 426, 2005.
- [11] M. Tumminello, F. Lillo, and R. N. Mantegna, "Correlation, hierarchies, and networks in financial markets," *Journal of Economic Behavior & Organization*, vol. 75, no. 1, pp. 40–58, 2010.
- [12] J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters, "Rotational invariant estimator for general noisy matrices," *arXiv preprint arXiv:1502.06736*, 2015.
- [13] B. Efron, "Bootstrap methods: another look at the jackknife," *The annals of Statistics*, pp. 1–26, 1979.

- [14] G. Carlsson and F. Mémoli, “Characterization, stability and convergence of hierarchical clustering methods,” *The Journal of Machine Learning Research*, vol. 11, pp. 1425–1470, 2010.
- [15] E. Levine and E. Domany, “Resampling method for unsupervised estimation of cluster validity,” *Neural computation*, vol. 13, no. 11, pp. 2573–2593, 2001.
- [16] S. Ben-David, U. Von Luxburg, and D. Pál, “A sober look at clustering stability,” in *Learning theory*. Springer, 2006, pp. 5–19.
- [17] O. Shamir and N. Tishby, “Cluster stability for finite samples.” in *NIPS*, 2007.
- [18] T. Di Matteo, F. Pozzi, and T. Aste, “The use of dynamical networks to detect the hierarchical organization of financial market sectors,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 73, no. 1, pp. 3–11, 2010.
- [19] S. Mayordomo, J. I. Peña, and E. S. Schwartz, “Are all credit default swap databases equal?” National Bureau of Economic Research, Tech. Rep., 2010.
- [20] G. Marti, P. Very, and P. Donnat, “Toward a generic representation of random variables for machine learning,” *arXiv preprint arXiv:1506.00976*, 2015.
- [21] M. Martens and S.-H. Poon, “Returns synchronization and daily correlation dynamics between international stock markets,” *Journal of Banking & Finance*, vol. 25, no. 10, pp. 1805–1827, 2001.
- [22] G. Marti, P. Donnat, F. Nielsen, and P. Very, “HCMapper: An interactive visualization tool to compare partition-based flat clustering extracted from pairs of dendrograms,” *arXiv preprint arXiv:1507.08137*, 2015.