Method Article

Sphere-sphere intersection for investment portfolio diversification — A new data-driven cluster analysis

Michel Ferreira Cardia Haddad

University of Cambridge, United Kingdom

A B S T R A C T

Aiming at supporting the process of investment portfolio diversification by using a data-driven approach, the present methodological paper proposes a new cluster analysis, which compares publicly traded companies, mainly in times of high volatility (e.g. crisis times). The main goal of the proposed method is to provide a less arbitrary analysis to support financial investors to precisely measure the degree of similarity between equity stocks, unveiling equity market clustering patterns by applying analytic geometry solutions and calculating an overall clustering pattern indicator. Empirical results on synthetic data demonstrate either that the proposed method has conceptual superiority over traditional cluster analyses and its potential practical usefulness to asset allocation, portfolio strategy, asset pricing, among other related purposes. Finally, the outputs of the proposed cluster analysis are presented through an intuitive and easily understandable mathematical visualization.

- It is proposed a new method to calculate risk-similarity and clustering patterns.
- The method unveils clustering patterns through a data-driven process.
- Portfolio diversification can benefit from sphere-sphere intersection calculations.

© 2019 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

A R T I C L E  I N F O

Method name: Data-driven cluster-similarity analysis

Keywords: Cluster analysis, Similarity measure, Asset allocation, Risk analysis, Investment decisions

Article history: Received 1 February 2019; Accepted 20 May 2019; Available online 25 May 2019

E-mail address: mfch2@cam.ac.uk (M.F.C. Haddad).

https://doi.org/10.1016/j.mex.2019.05.025

2215-0161/© 2019 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
**Specifications Table**

| Subject Area:            | Social Sciences |
|--------------------------|-----------------|
| More specific subject area: | Econophysics    |
| Method name:             | Data-driven Cluster-similarity Analysis |
| Name and reference of original method: | Main original method names: k-means Clustering and Hierarchical Cluster Analysis (HCA) and clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), pp. 651–666 |
| Resource availability:   | N/A             |

**Method details**

**Introduction**

The greater frequency and impact of financial and economic crises requires the development of feasible methods that aim to clearly distinguish investment alternatives in a robust, consistent, and coherent manner [1–4]. The main goal of the proposed data-driven cluster-similarity risk method is to support financial investors in precisely measure the level of similarity between publicly listed companies through time - especially in turbulent periods, showing the outputs through a clear graphical representation as a by-product of the analysis. As depicted in the Graphical Abstract, the method is performed through three subsequent steps, in which each asset is represented by a correspondent 2-sphere (i.e. ordinary three-dimensional sphere) in the three-dimensional Euclidean space ($\mathbb{R}^3$).

In step one is performed an analysis over the trajectory of each equity stock through time as well as the calculation of the variation of the individual risk factors – termed as $IR_{jt}$ – by finding the correspondent 2-sphere (simply sphere onwards) radius. Subsequently, in step two the spatial approximation between assets/spheres is visualized and the intersection volume – termed as $CR_{jt}$ – between every pair of spheres in the sample is calculated using analytical geometry. Finally, in step three either the individual spherical volumes and the intersection volumes calculated in the previous steps are used as input values placed into an overall clustering pattern indicator bounded from zero to one, which is a proxy measure designed to meaningfully assess the level of shared risk between all $q$ stocks in a sample based on their level of similarity at any particular date as well as through time $t$.

Most of the labour involved in the proposed method occurs in the first two steps, which refer to geometrical calculations and computations involving spheres and intersection volumes in the three-dimensional Euclidean space. However, the labour required is much compensated by the precise outputs generated, which results in consistent estimates as well as clear mathematical visualization of the calculations and analyses performed. The motivations to create and propose this non-hierarchical clustering method are based on an attempt to tackle some relevant problems and challenges often reported in the cluster analysis literature, which traditional clustering methods – such as $k$-means clustering and hierarchical cluster analysis (HCA) – do not address properly [5,6], such as: (i) setting the number of clusters in a non-arbitrary manner (i.e. through a data-driven process); (ii) possibility of a dataset having zero cluster; and (iii) allocating outliers to no cluster.

Although the proposed method was primarily designed to work with quantitative continuous variables as input data, it is possible to use virtually any type of data (e.g. nominal, ordinal, binary) to be placed into any of the three axes in $\mathbb{R}^3$. An important characteristic of using such non-continuous variables is that they can be represented as points in a multi-dimensional space, allowing for the visualization of the relationships between different datasets.

---

1. Both terms (i.e. asset and sphere) will be used interchangeably throughout this paper.
2. Regarding the index notation adopted in this paper, in an attempt to be as neat as possible, throughout the paper the indexes related to each of the sample subjects (e.g. $i$, $j$) as well as time (i.e. $t$) are separated by a vertical bar - i.e. |. Therefore, in the present paper, this symbol does not mean conditional probability, as commonly used in Bayesian statistics.
variables is that the spatial distance between each value in the axis must be equally spaced. Therefore, a relevant caveat of using such variables is that this equal spatial distance between each value needs to be set arbitrarily by the analyst. This arbitrary setting can potentially distort the analysis performed, by allowing the analyst to manipulate the outputs – for instance, by setting a very large spatial distance between each value in the axis in order to decrease the number of spherical volume intersections and, consequently, artificially reducing the number of clusters in a particular sample, which is extremely undesirable. This problem is avoided once input data are based on continuous variables since the spatial distance between each value in the axis would be determined solely by the data itself (i.e. data-driven), following a predetermined standard scale (e.g. company revenue, inflation rate, exchange rate).

Following this introduction, the paper is divided into five sections and it proceeds as follows. Sections Asset Individual Trajectory (Step 1), Spatial Approximation between Assets (Step 2) and Clustering Pattern Indicator (Step 3) explore and detail the first, second, and third steps of the proposed method, respectively. Section Experimental Outputs and Results provides two case studies and a brief discussion over the results. Finally, the last section concludes and suggests extensions for future research.

Step 1: asset individual trajectory

The proposed analysis starts by verifying the spatial trajectory of each stock (based on their respective axial variables) as well as the contemporaneous progress of their respective individual risk factor (based on their respective spherical volume). There are four values to be used as data input in the proposed analysis: the first, second, and third values are reflected on the x-axis, y-axis, and z-axis, respectively; and the fourth value is reflected in the volume of each sphere (i.e. $V_{ijt}$ or $IRF_{ijt}$).

Each of these four values (i.e. $x_{ijt}$, $y_{ijt}$, $z_{ijt}$, and $V_{ijt}$) is based on the rate of variation (from $t - 1$ to $t$) of four real-world variables (i.e. $s_{ijt}$, $k_{ijt}$, $o_{ijt}$, and $v_{ijt}$), respectively. The rate of variation of each of the four real-world variables to be placed into each of the three axes as well as reflected in the sphere volume are calculated as follows [7]:

\[
\begin{align*}
x_{ijt} &= \frac{s_{ijt} - s_{ijt-1}}{s_{ijt}}, & \forall x_{ijt}, s_{ijt} \in \mathbb{R} \\
y_{ijt} &= \frac{k_{ijt} - k_{ijt-1}}{k_{ijt}}, & \forall y_{ijt}, k_{ijt} \in \mathbb{R} \\
z_{ijt} &= \frac{o_{ijt} - o_{ijt-1}}{o_{ijt}}, & \forall z_{ijt}, o_{ijt} \in \mathbb{R} \\
V_{ijt} &= \frac{v_{ijt} - v_{ijt-1}}{v_{ijt-1}}, & \forall V_{ijt} \in \mathbb{R}^+, \forall v_{ijt} \in \mathbb{R}
\end{align*}
\]

Where $q = (1, 2, \ldots, i, j, \ldots, Q - 1, Q)$ and $t = (1, 2, \ldots, T - 1, T)$, which refer to the number of subjects and time span of the sample, respectively. Throughout this paper and in distinct contexts, the subjects $i$ and $j$ are frequently used to exemplify two generic sample subjects. Worth noting that the fourth variable is in absolute value, which reflects the volume of the respective sphere in $\mathbb{R}^3$. Therefore, the time series placed into the three axes as well as the one used as the spherical volume are sequenced as follows:

\[
\begin{align*}
x_{ijt} &= \{x_{1i1}, \ldots, x_{1iT}; x_{2i1}, \ldots, x_{2iT}; \ldots; x_{Qi1}, \ldots, x_{QiT}\} \\
y_{ijt} &= \{y_{1i1}, \ldots, y_{1iT}; y_{2i1}, \ldots, y_{2iT}; \ldots; y_{Qi1}, \ldots, y_{QiT}\} \\
z_{ijt} &= \{z_{1i1}, \ldots, z_{1iT}; z_{2i1}, \ldots, z_{2iT}; \ldots; z_{Qi1}, \ldots, z_{QiT}\}, & \text{and} \\
IRF_{ijt} := V_{ijt} &= \{V_{1i1}, \ldots, V_{1iT}; V_{2i1}, \ldots, V_{2iT}; \ldots; V_{Qi1}, \ldots, V_{QiT}\}
\end{align*}
\]
Each sphere represents a different stock within a particular industry in the three-dimensional Euclidean space. For instance, in the automotive industry would be reasonable to compared stock indicators (i.e. data input) from companies such as Fiat, Ford, General Motors, Toyota, and Volkswagen; in the technology industry, would be analysed data from companies such as Alphabet/Google, Amazon, Apple, Facebook, and Microsoft; among many more possibilities.

The volume of each sphere is analytically calculated and graphically depicted in $\mathbb{R}^3$. The spherical volume aims to reflect a variable that is recognised as the best proxy available to measure the level of risk of a particular data set. This variable should be relevant and meaningful according to the point of view of the investor/analyst who is performing the analysis. Moreover, this risk proxy should account for industry context and idiosyncrasies (e.g. levels of expected stock price return, equity-to-debt ratio, market capitalization, or any other well justified indicator). As illustrated in Fig. 1, this risk proxy varies dynamically according to each different time $t$ as well as distinct stock in the sample. Earnings per share (EPS), price to earnings (P/E) ratio, price to book value (P/B) ratio, and dividend yield are just a few examples, among a wide range of possibilities, of real-world economic variables and/or financial indicators to be placed at any of three axes in $\mathbb{R}^3$ (Fig. 1).

As applicable to any quantitative analysis, it is of utmost importance that the analyst selects coherently and justifies properly each variable to be used as input data to perform the proposed method, specially the fourth and most important variable to be reflected in the volume of the respective sphere (i.e. IRFi$_{ijt}$). The analysis over the question on how to choose the most appropriate variables is out of the scope of this paper, which can be based on a series of factors, such as individual investment preferences, risk tolerance, market consensus, among many others.

The reason behind the analysis being performed in the three-dimensional space $\mathbb{R}^3$ relies on the fact that, compared to the two-dimensional space $\mathbb{R}^2$, one additional variable is included in the analysis. This additional variable enriches the analysis by allowing more information being added as input data. Moreover, although feasible, the reason to not propose the present method in an $n$-dimensional Euclidean space $\mathbb{R}^n$ relies on the fact that, on one hand the analyst can use $n$ variables instead of three, which is desirable. On the other hand, however, in $\mathbb{R}^n$ there would not be a consistent graphical visualisation of the outputs.

**Asset individual risk factor: IRFi$_{ijt}$**

As depicted in Fig. 2, each equity stock is represented by its respective sphere in $\mathbb{R}^3$ and their individual volume is given by a real-world economic or financial variable, termed as IRFi$_{ijt}$, which is calculated as follows:

$$IRFi_{ijt} := V_{ijt} = r_{ijt}^3 \frac{\pi^{3/2}}{\Gamma\left(\frac{3}{2} + 1\right)} = r_{ijt}^3 \frac{4}{3}\pi, \quad IRFi_{ijt} \in \mathbb{R}^+, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}^+$$

![Fig. 1. Visualisation of spatial trajectories of stocks $S_{ij}$ and $S_{jt}$ in $\mathbb{R}^3$ from $t = 1$ (left) to $t = 2$ (right).](image)
Where $r_{ijt}$ is the time-varying radius of the sphere, $\pi$ is the well-known ratio of the sphere’s circumference to its diameter, and $\Gamma$ is the gamma function; all variables refer to stock $S_{ijt}$ at time $t$. The formulas in Eq. (2) refer to the classical formulas to calculate the volume of a sphere in a three-dimensional Euclidean space ($\mathbb{R}^3$) and it was adapted to the present paper in order to incorporate the temporal variation reflected in the volume as well as the time-varying radius of the respective sphere. In addition, in the case one wants to check the accuracy of Eq. (2) vis-à-vis a particular real-world data used as input in the proposed analysis, one can examine the accuracy of the transcendental number $\pi$, which surely must represent the same precise spherical proportion for any subject as well as in every date $t$ as per below:

$$\frac{c_{ijt}}{D_{ijt}} = \frac{C_{ijt}}{2r_{ijt}} \approx \pi \approx 3.14159265358979$$  \hspace{1cm} (3)

Where $c_{ijt}$ is the circumference and $D_{ijt}$ is the diameter of the sphere, both representing stock $S_{ijt}$ at time $t$.

**Time-varying radius: $r_{ijt}$**

As $\pi$ and $\Gamma$ are constants (and, therefore, not a function of time) and as $IRF_{ijt}$ is known in advance (because it refers to a given real-world data value used as the main risk proxy in the proposed analysis).
analysis), in fact the only unknown variable in Eq. (2) is the time-varying radius \( r_{ijt} \), which can be found as follows:

\[
\frac{V_{ijt}}{r_{ijt}^3} = \frac{4}{3} \pi \tag{4}
\]

Replacing the sphere volume \( V_{ijt} \) by \( IRF_{ijt} \) from Eq. (4) and solving for \( r_{ijt} \) gives:

\[
r_{ijt} = \left( \frac{1}{\sqrt[3]{IRF_{ijt}^{-1}}} \right) \frac{4}{3} \pi \tag{5}
\]

Thus, to the extent that time \( t \) passes and new events succeed (e.g. news in the media, communications made by governments, entry of new competitors, CEO retirement) stocks (represented by spheres) perform a spatial trajectory as well as experience spherical volume variation through time and eventually interact between themselves, reflecting real-world interactions in \( \mathbb{R}^3 \). For instance, if a company has one of its indicators (e.g. ROI, reflected in the \( x \)-axis) considered as the respective industry benchmarking (e.g. Apple within the technology industry, Exxon Mobil in the oil and gas sector), then its competitors tend to adopt similar management decisions and take action in an attempt to close the gap related to its own indicator (in this example, ROI) compared to the industry benchmarking one. As a consequence of that competition move, indicators of each stock in the sample tend to become more similar through time and, therefore, closer between themselves in \( \mathbb{R}^3 \).

**Step 2: spatial approximation between assets**

After calculating all asset individual risk factors for each stock in the sample in step 1, then it is possible to proceed to calculate as well as perform a visual inspection in \( \mathbb{R}^3 \) to confirm either if there is no risk volume sharing situation between stocks (i.e. case in which there are only trivial intersections between pairs of spheres in \( \mathbb{R}^3 \)) or if risk volume sharing between one or more pairs of spheres has

![Fig. 3. Non-risk volume sharing situation between stocks \( S_{ijt} \) and \( S_{jkt} \) in \( \mathbb{R}^3 \) at time \( t \), represented by a trivial intersection in which \( S_{ijt} \cap S_{jkt} = \emptyset \).](image-url)
effectively occurred (i.e. case in which there is at least one non-trivial intersection between spheres in \(\mathbb{R}^3\), as depicted in Figs. 3 and 4).

The measurement of the individual risk factor \(IRF_{i|t}\) (i.e. risk proxy of stock \(S_{i|t}\) or its spherical volume regardless of any intersection at a particular date \(t\)) is found in step 1 by calculating the individual volume of the sphere that represents each stock at each particular point in time \(t\). If there is an intersection between stocks \(S_{i|t}\) and \(S_{j|t}\) in \(\mathbb{R}^3\) then it is possible to calculate the respective shared value between those two stocks. This shared volume (i.e. non-trivial intersection) occurs due to (i) greater similarity in the spatial trajectory and proximity of the indicators between stocks; (ii) increase in the risk perception proxy measure of each stock, reflected in a greater sphere volume; or (iii) the combined performance of items (i) and (ii).

**Asset common risk factor: \(CRF_{i|j|t}\)**

The shared volume between stocks \(S_{i|t}\) and \(S_{j|t}\) is termed as the common risk factor \(CRF_{i|j|t}\) due to the fact that this value is attributed to individual risk factors \(IRF_{i|t}\) and \(IRF_{j|t}\) that are shared between the pair of stocks \(S_{i|t}\) and \(S_{j|t}\), as follows:

\[
CRF_{i|j|t} := S_{i|t} \cap S_{j|t}
\]
Analytic geometry is the obvious path to calculate the precise volume of two spheres intersecting in \( \mathbb{R}^3 \), as graphically depicted in Fig. 5. There is a non-trivial volume intersection between two spheres \( S_{ij} \) and \( S_{ji} \) in the case that, if and only if:

\[
\| C_{ij} - C_{ji} \| < r_{ij} + r_{ji}, \quad S_{ij} \cap S_{ji} \neq \emptyset
\]

(7)

Where \( C_{ij} \) and \( C_{ji} \) are the centres of spheres \( S_{ij} \) and \( S_{ji} \), respectively, and \( \| \cdot \| \) refers to the Euclidean norm.

**Definition 1.** The intersection of two spheres is the circumference of a circle whose plane is perpendicular to the line joining the centres of the surfaces and whose centre is in that line [8].

The spheres \( S_{ij} \) and \( S_{ji} \), of radii \( r_{ij} \) and \( r_{ji} \) are placed at time \( t \) and centred at \( C_{ij}(0,0,0) \) as well as \( C_{ji}(d_{ji},0,0) \) in \( \mathbb{R}^3 \), respectively. The calculation of the sphere-sphere volume intersection is similar to the circle-circle area intersection, which is in accordance with **Definition 1** and which plays an important role in the sphere-sphere volume intersection calculation. Below is detailed the steps of the analytical calculation of a non-trivial intersection between the pair composed by spheres \( S_{ij} \) and \( S_{ji} \) [9–12].

The equation of sphere \( S_{ij} \) is defined in Eq. (8) and the equation of sphere \( S_{ji} \) is defined in Eq. (9), as follows:

\[
r_{ij}^2 = x_{ij}^2 + y_{ij}^2 + z_{ij}^2
\]

(8)

\[
r_{ji}^2 = (x_{ij} - d_{ji})^2 + y_{ij}^2 + z_{ij}^2
\]

(9)

Combining Eq. (8) and Eq. (9) gives:

\[
r_{ji}^2 = (x_{ij} - d_{ji})^2 + (r_{ij}^2 - x_{ij}^2)
\]

(10)

Multiplying through and rearranging Eq. (10) result in the following:

\[
x_{ij}^2 - 2d_{ji}x_{ij} + d_{ji}^2 + x_{ij}^2 - r_{ji}^2 = r_{ij}^2 - r_{it}^2
\]

(11)

Solving Eq. (11) for \( x_{ij} \) gives:

\[
x_{ij} = \frac{d_{ji}^2 - r_{ji}^2 + r_{it}^2}{2d_{ji}}
\]

(12)
The intersection between both spheres refers to a curve lying in a plane parallel to the \(y_{ij} \), \(z_{ij} \)-plane at a single \(x_{ij} \)-coordinate, which plugging it back into Eq. (8) yields the following:

\[
y^2_{ij} + z^2_{ij} = r^2_{ij} - x^2_{ij}
\]  

(13)

\[
y^2_{ij} + z^2_{ij} = r^2_{ij} - \left(\frac{d^2_{ij} - r^2_{ij} + r^2_{ij} - r^2_{ij}}{2d_{ij}}\right)^2
\]  

(14)

\[
y^2_{ij} + z^2_{ij} = \frac{4d^2_{ij} + 2r^2_{ij} - \left(\frac{d^2_{ij} - r^2_{ij} + r^2_{ij} - r^2_{ij}}{2d_{ij}}\right)^2}{4d^2_{ij}}
\]  

(15)

Which refers to a circle with the following radius \(r_{\lambda ij} \):

\[
r_{\lambda ij} = \frac{1}{2d_{ij}} \sqrt{4d^2_{ij} + 2r^2_{ij} - \left(\frac{d^2_{ij} - r^2_{ij} + r^2_{ij} - r^2_{ij}}{2d_{ij}}\right)^2}
\]  

(16)

\[
r_{\lambda ij} = \frac{1}{2d_{ij}} \sqrt{\left(-d_{ij} + r_{ij} - r_{ij}\right)\left(-d_{ij} + r_{ij} + r_{ij}\right)\left(-d_{ij} + r_{ij} + r_{ij}\right)\left(d_{ij} + r_{ij} + r_{ij}\right)}
\]  

(17)

Thus, the volume of the three-dimensional common lens of spheres \(S_{ij} \) and \(S_{j} \) can be found by adding the respective spherical caps. The distances from the centres of spheres \(S_{ij} \) and \(S_{j} \) to the bases of the respective caps are given by:

\[
d_{\alpha ij} = x_{ij}
\]  

(18)

\[
d_{\beta ij} = d_{ij} - x_{ij}
\]  

(19)

The heights of each of the spherical caps are then calculated as follows:

\[
h_{ij} = x_{ij} - d_{\alpha ij} = \frac{\left(r_{ij} - r_{ij} + d_{ij}\right)\left(r_{ij} + r_{ij} - d_{ij}\right)}{2d_{ij}}
\]  

(20)

\[
h_{ij} = r_{ij} - d_{\beta ij} = \frac{\left(r_{ij} - r_{ij} + d_{ij}\right)\left(r_{ij} + r_{ij} - d_{ij}\right)}{2d_{ij}}
\]  

(21)

The volume of a spherical cap of height \(h_{\lambda} \) for a sphere of radius \(r_{\lambda} \) is:

\[
V_{\text{cap}, \lambda} \left(r_{\lambda}, h_{\lambda}\right) = \frac{1}{3} \pi h^2_{\lambda} \left(3r_{\lambda} - h_{\lambda}\right)
\]  

(22)

Therefore, in order to specifically find the \(C_{R i j} \), it is necessary to sum both spherical caps, as follows:

\[
C_{R i j} = V_{\text{cap}, \alpha} \left(r_{ij}, h_{ij}\right) + V_{\text{cap}, \lambda} \left(r_{ij}, h_{ij}\right)
\]  

(23)

\[
C_{R i j} = \frac{\pi \left(r_{ij} + r_{ij} - d_{ij}\right)^2 \left(d^2_{ij} + 2d_{ij}r_{ij} - 3r^2_{ij} + 2d_{ij}r_{ij} + 6r_{ij}r_{ij} - 3r^2_{ij}\right)}{12d_{ij}}
\]  

(24)
Where $V_{cap_i}$ and $V_{cap_j}$ refer to the volume of the spherical caps of spheres $S_{i,t}$ and $S_{j,t}$, respectively. In the case that $d_{ij} = r_{ij} + r_{ij}$, the expression above gives $CRF_{ij,t} = 0$ as one would expect.

This section considers only the intersection between the two spheres $S_{i,t}$ and $S_{j,t}$ in $\mathbb{R}^3$ at a particular point in time $t$. However, depending on the number of subjects $q$ included in the sample, it would be necessary to consider more than two spheres, performing intersection calculation between $q$ spheres (representing the $q$ stocks in the respective sample) at time $t$. For samples with $q > 2$, computational geometry techniques would be a suitable alternative to reach reasonably approximate spherical volume intersection results through numerical analysis.\(^3\)

**Asset idiosyncratic risk through symmetric difference: $IR_{ij,t}$**

Through symmetric difference one can easily find the idiosyncratic risk measure of stock $S_{i,t}$, termed as $IR_{ij,t}$, which refers to a risk value attributed only to individual factors of each stock at a particular time $t$. The calculation of this type of risk is performed to provide the remaining volume of the sphere that has no intersection with any other stock in the sample (i.e. total volume of the sphere representing the risk information carried by stock $S_{i,t}$ subtracted by the intersecting volume with other stock’s spheres), as follows:

$$IR_{ij,t} := IRF_{ij,t} - (IRF_{ij,t} \cap IRF_{ij,t})$$

(25)

The symmetric difference of $IRF_{ij,t}$ and $IRF_{ij,t}$ is associative as well as commutative, and can be alternatively denoted using the following notation [13,14]:

$$IRF_{ij,t} \triangle IRF_{ij,t} = IRF_{ij,t} \triangle IRF_{ij,t}$$

(26)

---

\(^3\) For further information on numerical analysis techniques to compute spheres intersection in three-dimensional spaces, see Strobl et al. [15] and George et al. [16].
The concept of $IR_{ijt}$ can also be interpreted as follows: although two publicly traded companies are experiencing problems at the same time $t$, by analysing their respective stock indicators used as input data in the proposed method, those two companies are potentially experiencing distinct problems and that is the reason that stocks $S_{ijt}$ and $S_{jkt}$ have very low or zero volume risk sharing at a particular point in time. Fig. 6 depicts the concept of the idiosyncratic risk $IR_{ijt}$.

Finally, in the case that $IR_{ijt} \cap IR_{jkt} = 0$, it can be interpreted that the overall risk of a particular stock $S_{ijt}$ is the result of individual (i.e. idiosyncratic) problems that are distinct, or even very distinct, from the problems being experienced by stock $S_{jkt}$ and, therefore, both stocks have zero volume risk sharing. Worth noting that care must be taken in order to read the results based on intersecting spherical volume values and then interpreting as well as translating them properly into meaningful similarity and risk measures that may support the decision-making process.

Step 3: clustering pattern indicator

In the last step, the values found in the two previous steps are used as input to an overall clustering pattern indicator, termed as $\mathcal{R}_t$. This indicator aims to measure how similar (or dissimilar) are sample subjects (e.g. stocks) at each point in time.

Times series cluster ratio of $CRF_{ijt}$ over $IR_{ijt}$ and $IR_{jkt}$: $\mathcal{R}_t$

After having found the individual risk factors $IR_{ijt}$ and $IR_{jkt}$ as well as the common risk factor $CRF_{ijt}$ between every pair of distinct spheres in the sample – except for intersections between the sphere with itself, such as $V_{1;i} \cap V_{1;i}$, which obviously yields $V_{1;i}$ – the respective intersecting volumes are placed in the numerator and the sum of individual volumes of all spheres are placed in the denominator of the following time series cluster ratio indicator:

$$\mathcal{R}_t := \frac{[IR_{1;i} \cap IR_{2;i}] + \ldots + [IR_{Q;i} \cap IR_{Q;i}]}{IR_{1;i} + IR_{2;i} + \ldots + IR_{Q;i} + IR_{Q;i}}$$

(27)

Which can alternatively be written as:

$$\mathcal{R}_t := \frac{(CRF_{1;i}) + \ldots + (CRF_{Q;i})}{IR_{1;i} + IR_{2;i} + \ldots + IR_{Q;i} + IR_{Q;i}}$$

(28)

Where

$$0 \leq \mathcal{R}_t \leq 1, \ \forall \mathcal{R}_t \in \mathbb{Q}^+$$

$$0 \leq IR_{1;i}, \ldots, IR_{Q;i} < +\infty, \ \forall IR_{1;i}, \ldots, IR_{Q;i} \in \mathbb{R}^+$$

$$0 \leq (CRF_{1;i}) + \ldots + (CRF_{Q;i}) \leq (IR_{1;i} + \ldots + IR_{Q;i}), \ (IR_{1;i} + \ldots + IR_{Q;i}) \in \mathbb{R}^+$$

In the case of the series $IR_{1;i}, \ldots, IR_{Q;i}$, the lower limit (i.e. zero) refers not only to the usual concept of an almost infinitesimal value, but also to an actual numerical possibility of being zero itself. On the other hand, the upper limit (i.e. $+\infty$) obviously does not refer to an actual value, but instead it is a mathematical concept that conveys the idea of an extremely large, however unreachable, number which in this case would reflect a very large spherical volume.

In the case in which there is a trivial intersection (i.e. $S_{ijt} \cap S_{jkt} = \emptyset$) the numerator would be zero, which would result in $\mathcal{R}_t = 0$. Conversely, in the opposite extreme case in which all spheres in the sample are placed in the exact same three coordinates and have the precise same volume, then $\mathcal{R}_t = 1$. Thus, on one hand, the closer $\mathcal{R}_t$ is to one, the more similar and grouped the stocks are in such a given sample. On the other hand, the closer $\mathcal{R}_t$ gets to zero, the more dissimilar and separated apart are the
stocks from each other. In analyses performed using real-world variables as data input, it is expected that most results should lie within those two extreme values that form the closed interval \([0, 1]\) as well as would be, although not impossible, extremely rare and unlikely to reach the precise upper bound value of one.

**Experimental outputs and results**

This section contains back-of-the-envelope calculations of two typical cases using synthetic data. In the first one it is considered a static case study in order to discuss what would be explored in a data set only in a particular point in time. The second case consists of a dynamic study of clustering patterns through time.

**Static case study**

Consider a financial investor who needs to build a portfolio limited to two distinct stocks amongst only four possible alternatives available in the equity market. Therefore, there are four stocks \(\{S_{A,t}, S_{B,t}, S_{C,t}, S_{D,t}\}\) and the respective rate of variation from \(t - 1\) to \(t\) of the following four variables are used as input data in this case study: return on investment (ROI); earnings per share (EPS); debt-to-equity ratio (D/E); and stock price simple net return (in absolute value). Subsequently, the first three

![Diagram](image)

**Fig. 7.** Static case study with a sample based on four stocks (i.e. \(S_{A,t}, S_{B,t}, S_{C,t}, \) and \(S_{D,t}\)) at time \(t = 1\).
variables are placed into the x, y, z-axes, respectively and the fourth variable is reflected in the volume of each of the four spheres \( \{ \text{IRF}_A, \text{IRF}_B, \text{IRF}_C, \text{IRF}_D \} \), as depicted in Fig. 7.

The red arrows in Fig. 7 are based on which variations of each of the three axial variables – as detailed in the header of Table 1 – would be beneficial or detrimental according to an investor’s point of view. Therefore, considering a typical rational investor, in virtually all scenarios the greater the ROI (i.e. the further to the right towards the x-axis), the better; the greater the EPS (i.e. the further up towards the y-axis), the better; and, conversely, the lower the D/E (i.e. the greater the profundity level towards the z-axis when the azimuthal angle is at 180°), the better. Overall, a rational investor would seek to maximise the return as well as minimise the risk of the investment. Therefore, in terms of return, an investor would prefer to maximise the positive variation of ROI, EPS, and Stock Return while positive (i.e. financial gain); and, conversely, would prefer to maximise the negative variation of D/E and Stock Return while negative (i.e. financial loss). On the other hand, in terms of risk, such a typical investor would prefer the lowest level of variation in all of these four variables due to the fact that the lower the volatility, the higher the level of predictability of such asset, which would result in a lower level of investment risk.

The simulated values of each of the variables used as input in this example are detailed in Table 1. Based only on a visual inspection on Fig. 7 as well as information provided by Table 1, one can draw the following preliminary and elementary conclusions, according to a rational investor’s point of view:

- Stock \( S_{A1} \) is better off than stock \( S_{B1}, S_{D1}, \) or \( S_{C1} \);
- Stock \( S_{B1} \) is better off than stock \( S_{D1} \) or \( S_{C1} \);
- Stock \( S_{D1} \) is better off than stock \( S_{C1} \); and
- \( \text{IRF}_{A1} > \text{IRF}_{B1} > \text{IRF}_{D1} > \text{IRF}_{C1} \).

Worth mentioning that one of the main goals of the proposed method is to assess, as impartially as possible, competing asset investment alternatives by measuring the similarity level between each pair of assets in a sample. Ultimately, the proposed method does not judge any of its input variables as

| Stock | \( \Delta \text{ROI} \) \( (x_{A1}) \) | \( \Delta \text{EPS} \) \( (y_{A1}) \) | \( \Delta \text{D/E} \) \( (z_{A1}) \) | \( \Delta \text{Stock Return} \) \( (\text{IRF}_{A1} \text{ or } V_{A1}) \) |
|-------|-----------------|-----------------|-----------------|-----------------|
| \( S_{A1} \) | 0.14            | 0.17            | -0.12           | 0.08            |
| \( S_{B1} \) | 0.09            | 0.13            | -0.07           | 0.07            |
| \( S_{C1} \) | 0.02            | 0.03            | 0.04            | 0.03            |
| \( S_{D1} \) | 0.04            | 0.06            | -0.02           | 0.05            |

Table 2

Hypothetical common values (i.e. intersecting spherical volumes) between each pair of stocks in the sample.

| \( CRF_{ij} \text{[l]} \) | \( S_{A1} \) | \( S_{B1} \) | \( S_{C1} \) | \( S_{D1} \) |
|-----------------|------------|------------|------------|------------|
| \( S_{A1} \)    | 0.080      | 0.045      | 0          | 0          |
| \( S_{B1} \)    | 0.045      | 0.070      | 0          | 0          |
| \( S_{C1} \)    | 0          | 0.030      | 0.021      | 0.050      |
| \( S_{D1} \)    | 0          | 0          | 0.021      | 0.050      |

Note: The values on the main diagonal refer to intersections between the sphere with itself, which yields the respective individual spherical volume. Such values are not used in the subsequent calculations/analyses.
beneficial or detrimental for the overall performance of an asset portfolio. For instance, it may be the case that a conservative investor A would prefer a lower level of D/E (e.g., lower probability of the underlying company going bankrupt as a consequence of not being able to repay its debts to creditors), while an investor B, with an aggressive profile, would prefer a higher level of D/E (e.g., expecting a higher profit in the near future due to: (i) a higher gearing ratio, meaning more funding financing company's projects, which is expected to be transformed into higher profits, and (ii) a greater tax-shield). Therefore, as the proposed method has a data-driven approach, such subjective and qualitative judgment must be made only by the decision maker, who would potentially benefit from non-biased and impartial insights provided by the proposed method.

Subsequently, using data from Table 1 as input data in the proposed method, one can calculate the values of the two intersections involving the whole sample of four stocks, one of them between stocks $S_{A1}$ and $S_{B1}$ and the other one related to stocks $S_{C1}$ and $S_{D1}$. After the calculation of the respective intersection volumes, it is possible to unveil the outputs shown in Tables 2 and 3.

Therefore, a potential financial investor seeking for stock market investment diversification opportunities can draw insightful conclusions through a visual inspection on Fig. 7 and, most importantly, based on the simulated calculations provided by Tables 2 and 3 such as, but certainly not limited to:

- $S_{A1}$ or $S_{B1}$ does not interact with stock $S_{C1}$ or $S_{D1}$, and vice-versa;
- $S_{A1}$ is more similar to $S_{B1}$ in comparison with stocks $S_{C1}$ or $S_{D1}$;
- $S_{B1}$ is more similar to $S_{A1}$ in comparison with stocks $S_{C1}$ or $S_{D1}$;
- $S_{C1}$ is more similar to $S_{D1}$ in comparison with stocks $S_{A1}$ or $S_{B1}$;
- $S_{D1}$ is more similar to $S_{C1}$ in comparison with stocks $S_{A1}$ or $S_{B1}$;
- There are two data-driven clusters in this sample. Cluster 1 is composed by stocks $S_{A1}$ and $S_{B1}$ and the members of the Cluster 2 are stocks $S_{C1}$ and $S_{D1}$;
- Cluster 1 ($S_{A1}$ and $S_{B1}$) is slightly more homogenous than Cluster 2 ($S_{C1}$ and $S_{D1}$) due to the fact that Cluster 1 has a similarity factor mean of 0.60, compared to 0.53 of Cluster 2. The means of the similarity factors are calculated as follows:

\[
\text{Cluster 1} = \left( \frac{\text{CRF}_{A,B1} + \text{CRF}_{B,A1}}{\text{IRF}_{A1} + \text{IRF}_{B1}} \right) = \left( \frac{0.045 + 0.045}{0.08 + 0.07} \right) = 0.600
\]

\[
\text{Cluster 2} = \left( \frac{\text{CRF}_{C,D1} + \text{CRF}_{D,C1}}{\text{IRF}_{C1} + \text{IRF}_{D1}} \right) = \left( \frac{0.021 + 0.021}{0.03 + 0.05} \right) = 0.525
\]

Overall, in terms of the pairwise risk-return trade-off, on one hand as a pair of assets and in terms of return, the pair of stocks $S_{A1}$ and $S_{B1}$ overperforms the pair $S_{C1}$ and $S_{D1}$. On the other hand, for portfolio composition strategy and taking into account diversification purposes in order to decrease the portfolio risk by choosing stocks as distinct as possible within a given pair, a blended portfolio with a stock from Cluster 1 (i.e., $S_{A1}$ or $S_{B1}$) and a stock from Cluster 2 (i.e., $S_{C1}$ or $S_{D1}$) would result in a higher degree of dissimilarity and would be more suitable in the case the investor aims to build a portfolio as dissimilar as possible (i.e., higher level of diversification given a certain level of return).

**Dynamic case study**

The rationale behind $r_{x_t}$ (detailed in the section Clustering Pattern Indicator) is applied to a sample of two stocks ($S_{A1}$ and $S_{B1}$) in a monthly basis through a whole year ($T = 12$), being each time $t$ corresponding to the last trading day of each of the 12 months of the year of 2008, which includes periods during the Global Financial Crisis of 2007–2008 (GFC of 2007–08 hereinafter). The four variables in this hypothetical dynamic case study are: $\Delta\text{ROI}$, $\Delta\text{EPS}$, $\Delta\text{D/E}$; and $\Delta$ stock price simple net return (in absolute value); all variables varying from $t = 1$ to $12$. 
Table 3  
Calculated similarity factors between each pair of stocks in the sample.

| $CRF_{AB}[1]$ | $IRF_{A}[1]$ | $IRF_{B}[1]$ | $IRF_{C}[1]$ | $IRF_{D}[1]$ |
|---------------|-------------|-------------|-------------|-------------|
| $CRF_{AB}[1]$ | 0.563       | 0.643       | *           | *           |
| $CRF_{AC}[1]$ | 0           | 0           | 0           | 0           |
| $CRF_{AD}[1]$ | 0           | 0           | 0           | 0           |
| $CRF_{BC}[1]$ | 0.563       | 0.643       | *           | *           |
| $CRF_{BD}[1]$ | 0           | 0           | 0           | 0           |
| $CRF_{CD}[1]$ | 0           | 0           | 0           | 0           |
| $CRF_{DC}[1]$ | *           | *           | 0.700       | 0.420       |

Note: The asterisks (*) refer to not meaningful interactions of the ratio $CRF_{AB}/IRF_{A}$, although still possible to be mathematically calculated. For instance, $CRF_{AB}/IRF_{C}$, in which $S_{C1} \cap S_{A1} = \emptyset$ and $S_{C1} \cap S_{B1} = \emptyset$.

As shown in Table 4, the first value in the time series of $\mathcal{R}_t$ (i.e. $t = 1$, reflecting January 2008) is the same value from the previous case study (subsection Static Case Study) for each of the two stocks, which is calculated as follows:

$$\mathcal{R}_1 = \frac{2(\mathcal{I}RF_{A1} \cap \mathcal{I}RF_{B1})}{\mathcal{I}RF_{A1}^2 + \mathcal{I}RF_{B1}^2} = \frac{2(0.045)}{0.08 + 0.07} = 0.600$$

Eleven months later, in $t = T = 12$ (i.e. December 2008), $\mathcal{R}_t$ in the last date of the sample results in:

$$\mathcal{R}_{12} = \frac{2(\mathcal{I}RF_{A12} \cap \mathcal{I}RF_{B12})}{\mathcal{I}RF_{A12}^2 + \mathcal{I}RF_{B12}^2} = \frac{2(0.107)}{0.130 + 0.120} = 0.856$$

Worth noting that two data features determine the value of each time $t$ in the series $\mathcal{R}_t$: (i) the level of similarity between the volumes of stocks $S_{A_t}$ and $S_{B_t}$ (which are placed in the denominator of $\mathcal{R}_t$) – regardless the value of their intersecting volume, and (ii) the intersecting volume itself between both stocks (which is placed in the numerator of $\mathcal{R}_t$). The greater (i) and (ii), the closer $\mathcal{R}_t$ gets from its maximum value of one (i.e. the extreme case in which both stocks would be identical and, therefore, they would have a full non-trivial intersection).

Table 4 contains input values and, most importantly, calculated risk-similarity measures of stocks $S_{A_t}$ and $S_{B_t}$ in a monthly basis throughout the sample time span as well as basic sample statistics, including the first and second central moments of each variable, with emphasis on the $\mathcal{R}_t$ variable (rightmost column).

As depicted in Figs. 8 and 9, according to the $\mathcal{R}_t$ measure there was an overall increase in the similarity level between stocks $S_{A_t}$ and $S_{B_t}$ throughout the year of 2008. More specifically, there was an increase of 43% from the beginning ($\mathcal{R}_1 = 0.60$) to the end of the year ($\mathcal{R}_{12} = 0.86$), and this similarity level growth was much steeper from the mid of the year ($\mathcal{R}_6 = 0.27$) to the end of the year ($\mathcal{R}_{12} = 0.86$), resulting in an increase of 221% from June to December of 2008.

In summary, by reading these outputs one can conclude that the similarity level between both stocks experienced a relevant increase from the beginning to the end of the year of 2008, possibly, among other potential factors, as a consequence of negative events and media news related to the GFC of 2007–08, when both stocks became increasingly more similar between themselves.
Conceptual comparison with well established methods

Overall, the proposed method in this paper can be understood as a non-hierarchical hard cluster analysis. One of the main motivations to create and propose this novel cluster analysis is to address some relevant problems and challenges frequently reported in the cluster analysis literature. More specifically, the following three problems are not currently being treated properly by most traditional clustering methods: (i) setting the number of clusters in a non-arbitrary manner; (ii) possibility of a dataset having zero cluster; and (iii) allocating outliers to no cluster.

The proposed method aims to tackle these three clustering problems at once by adopting a distinct and unusual strategy: imposing an artificial boundary on each sample subject by representing them as spheres instead of points in the three-dimensional Euclidean space. This artificial boundary

| t         | IRF_{SA} | IRF_{SB} | IRF_{SA} + IRF_{SB} | IRF_{SA} \cap IRF_{SB} | K_t |
|-----------|----------|----------|---------------------|-------------------------|-----|
| 31-Jan-2008 | 0.080    | 0.070    | 0.150               | 0.045                   | 0.600 |
| 29-Feb-2008 | 0.060    | 0.020    | 0.080               | 0.017                   | 0.425 |
| 31-Mar-2008 | 0.070    | 0.030    | 0.100               | 0.017                   | 0.340 |
| 30-Apr-2008 | 0.050    | 0.060    | 0.110               | 0.021                   | 0.382 |
| 30-May-2008 | 0.080    | 0.100    | 0.180               | 0.027                   | 0.300 |
| 30-Jun-2008 | 0.040    | 0.050    | 0.090               | 0.012                   | 0.267 |
| 31-Jul-2008 | 0.110    | 0.120    | 0.230               | 0.041                   | 0.357 |
| 29-Aug-2008 | 0.080    | 0.060    | 0.140               | 0.033                   | 0.471 |
| 30-Sep-2008 | 0.170    | 0.210    | 0.380               | 0.142                   | 0.747 |
| 31-Oct-2008 | 0.130    | 0.190    | 0.320               | 0.125                   | 0.781 |
| 28-Nov-2008 | 0.150    | 0.160    | 0.310               | 0.133                   | 0.858 |
| 31-Dec-2008 | 0.130    | 0.120    | 0.250               | 0.107                   | 0.856 |

Mean (\bar{\mu}) | 0.096 | 0.099 | 0.195 | 0.060 | 0.532 |
Variance (\sigma^2) | 0.002 | 0.004 | 0.010 | 0.003 | 0.050 |
Std. Deviation (\sigma) | 0.041 | 0.062 | 0.101 | 0.051 | 0.224 |
Sample Size (n) | 12 | 12 | 12 | 12 | 12 |

Fig. 8. Line chart depicting the \( R_t \) time series and two major negative events during the GFC of 2007-08, from January 2008 to December 2008.
autonomously (i.e. through a data-driven process) determines the degree of similarity between each pair of subjects in the sample. As the proposed method is based on a data-driven approach, once the input data are included in the model, this method can potentially provide insightful results to the analyst, such as the following two simple examples: (i) in the case only one of the sample subjects is very distinct from all of its counterparts, then this particular subject should not have intersection volume with any other subject in the sample, being represented as an isolated sphere in $\mathbb{R}^3$ (i.e. outlier candidate); (ii) in the case every subject in the sample is very dissimilar between themselves, then one feasible result is the possibility of no sphere-sphere intersection, resulting in $R_t = 0$ (i.e. a sample in which there is no cluster being formed “naturally” by the input data).

The aforementioned brief description of improvements and output possibilities refer to an advantage over traditional cluster analyses, such as connectivity models (e.g. HCA) or centroid models (e.g. $k$-means clustering). Finally, an interesting and more reasonable empirical comparison would be against density models, such as DBSCAN (density-based spatial clustering of applications with noise) and/or OPTICS (ordering points to identify the clustering structure).

**Conclusion and future research**

Through a more consistent as well as less arbitrary data-driven new hard cluster method, the outputs aim to specifically support financial investors in their portfolio diversification and asset
reallocation strategy. The proposed method analyses similarity patterns between pairs of stocks as well as the market clustering structure through time.

The empirical cases based on hypothetical results shown in this paper illustrate how the proposed method can be potentially useful and relevant for financial investment purposes, more specifically for asset reallocation decision making and portfolio diversification strategy. Moreover, the proposed analysis is not limited to the traditional risk-return trade-off analysis, but rather uses a greater number of dimensions based on investor needs and preferences. This type of financial clustering pattern analysis may be useful mainly when equity stocks tend to become more similar (e.g. during crisis or high volatility times) and, therefore, more difficult to clearly distinguish investment alternatives by using traditional cluster methods available.

As suggestions for future research, a natural subsequent stage refers to the use of real-world variables as input data. In fact, the proposed method can potentially be applied to a wide range of fields of knowledge and research problems – as diverse as biology, medicine, geopolitics, engineering, international trade agreements, optimal M&A partnership, marketing, among many more, in order to compare similarity levels between sample subjects and, therefore, not being limited only to applications in finance and economics. In addition, the development of an algorithm (e.g. MatLab code/program) to perform all steps of the proposed method and the use of real-world economic variables as input data on this method are on-going projects. Needless to mention that insights based on outputs to be obtained by the application of the proposed method on distinct data sets remain an empirical question.

Acknowledgements

This paper benefited from stimulating conversations with Philip Arestis, Mardi Dungey, Andrew Harvey, Mohammad Hashem Pesaran, Jacob Rasmussen, John Glascock, John Howe, Neil Dodgson, and Alan Thompson, as well as two anonymous referees for their careful reading and insightful comments and suggestions. All errors remain exclusively the author’s responsibility. The present research was supported by Coordination for the Improvement of Higher Education Personnel of Brazil (CAPES) and The Cambridge Commonwealth, European & International Trust (under grant BEX 2220/15–6).

References

[1] T.G. Andersen, T. Bollerslev, F.X. Diebold, H. Ebens, The distribution of realized stock return volatility, J. Financ. Econ. 61 (1) (2001) 43–76.
[2] G.W. Schwert, Business cycles, financial crises, and stock volatility (September), Carnegie-Rochester Conference Series on Public Policy 31 (1989) 83–125.
[3] E.M. Reinhart, K.S. Rogoff, This Time Is Different: Eight Centuries of Financial Folly, Princeton University Press, 2009.
[4] P. Arestis, E. Karakitsos, Lessons from the ‘Great Recession’ for both theory and economic policy, Financial Stability in the Aftermath of the ‘Great Recession’, Palgrave Macmillan, London, 2013, pp. 164–192.
[5] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. 31 (8) (2010) 651–666.
[6] C. Hennig, M. Meila, Cluster analysis: an overview, Handbook of Cluster Analysis, CRC Press, Taylorand Francis Group, 2015, pp. 1–20.
[7] R.S. Tsay, Analysis of Financial Time Series, vol. 543, John Wiley & Sons, 2005.
[8] W.F. Kern, J.R. Bland, Solid Mensuration: With Proofs, J. Wiley & Sons, Incorporated, 1938.
[9] B. Alvord, The intersection of circles and the intersection of spheres, Am. J. Math. 5 (1) (1882) 25–44.
[10] N.A. Court, Four intersecting spheres, Am. Math. Mon. 67 (3) (1960) 241–248.
[11] V.N. Serezhkin, Y.N. Mikhailov, Y.A. Buslaev, The method of intersecting spheres for determination of coordination numbers of atoms in crystal structures, Russ. J. Inorg. Chem. 42 (12) (1997) 1871–1910.
[12] E.W. Weisstein, Sphere-Sphere Intersection, (2007) [accessed 2019]. Available at: http://mathworld.wolfram.com/Sphere-SphereIntersection.html.
[13] E.J. Borowski, J.M. Borwein, The HarperCollins Dictionary of Mathematics, HarperCollins, 1991.
[14] H. Alt, U. Fuchs, G. Rote, G. Weber, Matching convex shapes with respect to the symmetric difference, Algorithmica 21 (1) (1998) 89–103.
[15] S. Strobl, A. Formella, T. Pöschel, Exact calculation of the overlap volume of spheres and mesh elements, J. Comput. Phys. 311 (2016) 158–172.
[16] P.L. George, F. Hecht, E. Saltel, Automatic mesh generator with specified boundary, Comput. Methods Appl. Mech. Eng. 92 (3) (1991) 269–288.