TUCS

Peter Sarlin

Data and Dimension Reduction for Visual Financial Performance Analysis

Turku Centre *for* Computer Science

TUCS Technical Report
No 1049, May 2012

Data and Dimension Reduction for Visual Financial Performance Analysis

Peter Sarlin
Åbo Akademi University, Department of Information Technologies

TUCS Technical Report
No 1049, May 2012

Abstract

This paper assesses the suitability of data and dimension reduction methods, and data-dimension reduction (DDR) combinations, for visual financial performance analysis. Motivated by no comparable quantitative measure of all aspects of dimension reductions, this paper attempts to capture the suitability of methods for the task through a qualitative comparison and illustrative experiments. While the discussion deals with differences of DDR combinations in terms of their properties, the experiments illustrate their general applicability for financial performance analysis. The main conclusion is that topology-preserving DDR combinations with predefined grid shapes, such as the Self-Organizing Map, are ideal tools for this task. We illustrate advantages of these types of methods with a visual financial performance analysis of large European banks.

Keywords: financial performance analysis; visualization; data reduction; dimension reduction

TUCS Laboratory
Data Mining and Knowledge Management Laboratory

1. Introduction

The ongoing global financial crisis and European debt crisis has highlighted the importance of a thorough understanding of financial entities, be they economies, markets or institutions. Fortunately, rapid advances in information technology have enabled access to huge databases of financial information. Unfortunately, analyzing these data is not completely unproblematic. Except for incompleteness of data due to missing values and comparability issues due to reporting differences (e.g. Nobes, 2006), as well as outliers and skewed distributions (e.g. Deakin, 1976), the multidimensionality of the problem is a central challenge for comprehension. In the case of firm-level financial entities, performance can be measured along several subdimensions, such as asset and management quality, as well as capital adequacy, earnings performance and liquidity ratios. Factors that further complicate the assessment of these high-dimensional and big data are temporal and cross-sectional dependencies and relations. Generally, performance analysis of banks can be attempted with various aims, such as studies on determinants of entity performance (e.g. Zopounidis *et al.*, 1995), predictions of undesired entity-level events (e.g. Demyanyk and Hasan, 2010) and efficiency measurement of entities (e.g. Halkos and Salamouris, 2004). While a wide range of methods have been applied to fulfill these aims, they seldom focus on providing the end-user with representations of data in easily understandable formats. A visualization or abstraction of high-dimensional data supports the knowledge crystallization process by enabling utilization of the pattern recognition capabilities of the human brain for exploring enormous databases.

Data and dimension reduction techniques, and their combination for data-dimension reduction (DDR), obviously hold some promise for representing data in an easily understandable format. Data reductions provide overviews of data by compressing information, whereas dimension reductions provide low-dimensional overviews of similarity relations in data. While the quality of data reductions can be quantified by common evaluation measures like quantization error, assessing the superiority of one dimension reduction method over others with a quantitative measure is more difficult. Since the mid-20th century, the overload of available data has stimulated a soar in the development of dimension reduction methods with inherent differences. Most of the methods fall into those aiming at distance and topology preservation (for an overview, see Section 2 and Lee and Verleysen, 2007). The distance-preserving methods have their basis in those preserving spatial distances in the data set (e.g. the family of Multidimensional scaling (MDS) methods), while topology-preserving methods have their basis in those preserving neighborhood relations in the data set (e.g. Self-Organizing Map (SOM) (Kohonen, 1982)). However, most differences in the quality of dimension reductions, as all structural information can impossibly be preserved in a lower dimension, derive from variations in preserved similarity relations (i.e. objective function), such as pairwise distances or topological relationships. The performance, and choice of model specification, of one method can generally be motivated by its own quantitative quality measure. However, the relative goodness of different methods depends strongly on the correspondence between the particular measure and the objective function.

The large number of dimension reduction methods has obviously stimulated quality comparisons along different measures. However, despite many attempts, inconsistency of the comparisons has lead to no unanimity on the superiority of one method (e.g. Flexer, 1997; 2001; Venna and Kaski, 2001). This also verifies that the goodness of methods depends to a large extent on the correspondence between the measure and the objective function, and confirms that the quality measure is a user-specified parameter depending on the task at hand. While recent advances in unified measures for evaluating dimension reductions have included a parameter for the user to specify properties that are more important to be preserved (Lee and Verleysen, 2009; Lueks *et al*., 2011), quantitative measures still have difficulties in including qualitative differences in properties of methods, such as differences in flexibility for difficult data and the shape of the low-dimensional output. This motivates assessing the suitability of data and dimension reduction methods for a specific task from a qualitative perspective.

The main aim of this paper is to capture the most suitable methods for visual financial performance analysis according to the needs for the task. To date, the most popular method for the task has been the SOM, which is oftentimes asserted as an artifact of its simplicity and intuitive formulation (e.g. Trosset, 2008; Lee and Verleysen, 2007). Yet, being well-known or simple, while being an asset, is not a proper validation of relative goodness. Hence, we assess the suitability of three classical, or so-called first-generation, dimension reduction methods for financial performance analysis: metric MDS (Torgerson, 1952), Sammon's (1969) mapping and the SOM. Rather than being the most recent methods, the rationale for comparing these methods is to capture the suitability of well-known dimension reduction methods with inherently different aims: global and local distance preservation and topology preservation. For DDR, we test serial and parallel combinations of the projections with three data reduction or compression methods: vector quantization (VQ) (Linde *et al*., 1980), *k*-means clustering (MacQueen, 1967) and Ward's (1963) hierarchical clustering. The relative goodness of methods for financial performance analysis will first be discussed from a qualitative perspective, in particular with respect to the needs for this task. That is, building low-dimensional mappings from high-dimensional and big data that function as displays for additional information, be it individual data (e.g. time series of entities) or general structural properties of data (e.g. distance structures and densities). Specifically, we pinpoint this into four key issues: form of structure preservation, computational cost, flexibility for problematic data and shape of the output. Then, experiments on a dataset of annual financial ratios for European banks will be used to illustrate the general applicability of the DDR combinations for the task. Results of these comparisons are then projected to the second generation of dimension reduction methods for a final discussion on the superiority of methods for visual financial performance analysis.

The paper is structured as follows. In Section 2, we review the literature related to financial performance analysis and data and dimension reduction. While Section 3 briefly introduces the data and dimension reduction methods used in this paper, Section 4 discusses optimal DDR combinations for financial performance analysis and disentangles the needs and aims into concrete properties for measuring suitability. Section 5 compares data and dimension reduction methods for financial performance analysis by a qualitative discussion, presents the financial dataset and pre-processing framework and shows some illustrative experiments. Section 6 discusses critically the results and draws parallels with the second-generation of dimension reduction methods. Finally, Section 7 concludes by presenting key findings and suggestions for future work.

2. Related literature

This section reviews related literature from three distinct directions. We discuss the literature related to measuring financial performance of banks, we present a brief taxonomy of data and dimension reduction methods as well as a review of data reduction methods, and finally discuss comparisons of data and dimension reductions.

2.1. Financial performance analysis of banks

Measuring performance of banks is a common task and has been performed with a wide variety of methods and aims. Depending on the definition of "performance", different methods have been applied for assessing and measuring characteristics of banks. The complexity is further increased as one may also be interested in properties of entities across cross-sections and their evolution over time. We approach the literature from three directions. Thereafter, we give an outlook on visualizing bank-related data.

First, there is a broad literature on determinants of bank performance. These types of studies have mostly utilized high-dimensional data and conventional statistical techniques (e.g. Fraser *et al.*, 1974; Zopounidis *et al.*, 1995). Second, a large number of studies have created early warning models in attempts to predict future distress in bank performance of two forms: country-specific banking crises (e.g. Demirguc-Kunt and Detragiache, 2000) and bank-specific bank distress (e.g. González-Hermosillo, 1999; Männasoo and Mayes, 2009). Recently, some advances in predictive capabilities have been gained through non-parametric methods (Demyanyk and Hasan, 2010). Third, there is a field on its own on measuring the efficiency of individual banks. This has most commonly been performed with data envelopment analysis (Halkos and Salamouris, 2004), fuzzy set theory (Seçme *et al.*, 2009) and grey relation analysis (Kung and Wen, 2007). Even though there is no one method that is suited for every purpose and situation, the above methods do not even propose to be applicable for visual performance analysis.

Visualization of bank-related data concerns a wide range of approaches, such as network models for illustrating interconnectedness, extended plots for assessing raw data or their summaries and analytical methods. Following the proposition by Aigner *et al.* (2008) on a better integration of visual, analytical and user-centered methods, the focus herein is on visualization through analytical methods for data and dimension reduction. Dimension reduction methods, while not being highly common for visualization in financial analysis, have previously been employed for visual financial performance analysis. MDS and the SOM have been applied in illustrating the performance of banks on a low-dimensional display: MDS was used for illustrating the performance of Spanish banks (Ezzamel and Mar-Molinero, 1987) and the SOM was used for mapping close to European banks (Sarlin and Eklund, 2011). In addition, while there exist applications of both MDS (Mar-Molinero and Serrano-Cinca, 2001) and the SOM (e.g. Martín-del-Brío and Serrano-Cinca, 1993) for predicting failures of banks, the SOM has remained its popularity not only in the domain of bank analysis, but also in other financial areas.[1]

2.2. Data and dimension reduction methods

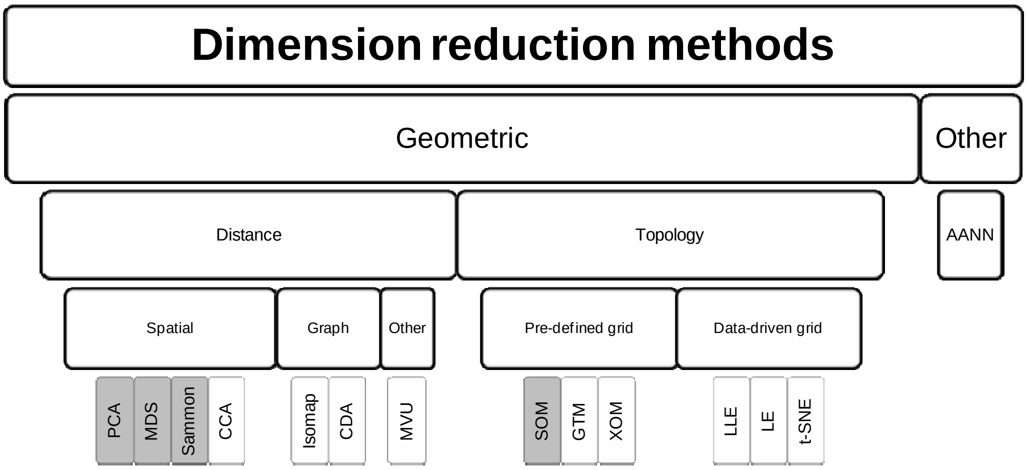
As the focus of this paper lies on dimension reductions, and data reductions are mainly used for enhancing the interpretation of dimension reductions, the focus of this review also lies on dimension reductions. The first dimension reduction methods date back to the early 20th century. However, only since the 1990s has there been a significant soar in the number of developed dimension reduction methods. We use that as a cutting point for dividing the methods into first and second-generation methods.

The first generation consists of the well-known classical methods that are still broadly used and accepted in a wide range of domains. Drawing upon the first introduced, but still commonly used, variance-preserving principal component analysis (PCA) (Pearson, 1901), an entire family of distance-preserving MDS-based methods have been developed. The MDS counterpart to PCA, classical metric MDS, that attempts to preserve pairwise distances was proposed by Young and Householder (1938) and Torgerson (1952). Non-linear versions are the first introduced non-metric MDS by Shepard (1962) and Kruskal (1964) and the later developed Sammon's (1969) mapping. The topology-preserving family of methods was launched through the introduction of the SOM (Kohonen, 1982). The SOM differs by reducing both dimensions and data through a neighborhood-preserving vector quantification. The second generation is a less homogeneous group of methods ranging from spectral techniques to graph embedding. A soar in developed methods at the turn of the century lead to several innovative techniques, such as Curvilinear Component Analysis (CCA) and Curvilinear Distance Analysis (CDA) (Demartines and Hérault, 1997), Generative Topographic Mapping (GTM) (Bishop *et al*., 1998a), Locally Linear Embedding (LLE) (Roweis and Saul, 2000), Isomap (Tenenbaum *et al*., 2000), Laplacian Eigenmaps (LE) (Belkin and Niyogi, 2001) and Maximum Variance Unfolding (MVU) (Weinberger and Saul, 2005). Some more recent methods are, for instance, t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten and Hinton, 2008) and Exploration Observation Machine (XOM) (Wismüller, 2009).

In addition to two generations, dimension reduction methods can also be illustrated in a tree-structured taxonomy. The tree-structure in Figure 1 is a non-exhaustive taxonomy of dimension reduction methods based upon that in Lee and Verleysen (2007, p. 234). While the focus herein is on methods based upon geometrical concepts, there exists also other methods, such as those based upon auto-associative neural networks (AANNs). The tree structure ends with some exemplifying methods, where first-generation methods are differentiated from second-generation methods through a grey background. Methods can roughly be divided into those

[1] The SOM has been popular for financial benchmarking and performance analysis in other financial domains as well. Some examples of financial application areas are the pulp and paper industry (Back *et al*. 1998; Eklund *et al*., 2004) and macro-financial surveillance (Sarlin and Peltonen, 2011)

aiming at distance and topology preservation. The distance-preserving methods can still be divided into different distances, such as spatial (e.g. PCA, MDS, Sammon's mapping and CCA), graph (e.g. Isomap and CDA) and other (e.g. MVU). Topology-preserving methods can be divided into those with a pre-defined grid shape (e.g. SOM, GTM and XOM) and those without (e.g. LLE, LE and t-SNE).



Notes: The figure represents a non-exhaustive taxonomy of dimension reduction methods adapted from Lee and Verleysen (2007). The lowest level associates methods to their families, where a gray background indicates first-generation methods and white second generation. Acronyms: Auto-Associative Neural Network (AANN), Principal Component Analysis (PCA), Multidimensional Scaling (MDS), Curvilinear Component Analysis (CCA), Curvilinear Distance Analysis (CDA), Maximum Variance Unfolding (MVU), Self-Organizing Map (SOM), Generative Topographic Mapping (GTM), Exploration Observation Machine (XOM), Locally Linear Embedding (LLE), Laplacian Eigenmaps (LE) and t-distributed Stochastic Neighbor Embedding (t-SNE).

Figure 1. A taxonomy of dimension reduction methods

Although there is no common taxonomy of data reduction methods, several properties can be used for differentiating between methods: soft vs. hard clustering, hierarchical vs. non-hierarchical methods and monothetic vs. polythetic goals, for instance. While soft clustering reduces data by assigning them to each cluster to a certain degree, hard clustering either assigns data to a cluster or not. Hierarchical methods (e.g. Ward's method (Ward, 1963)) produce a taxonomy of cluster structures, in which small child clusters are also nested within larger parent clusters, and may be divided into agglomerative (bottom-up) and divisive (top-down) approaches. Non-hierarchical methods approach data reduction from numerous different angles, and may roughly be divided into centroid-based (e.g. *k*-means (MacQueen, 1967) and Vector Quantization (Linde *et al*., 1980)), distribution-based (e.g. Expectation-maximization algorithm (Dempster *et al*., 1977)) and density-based clustering (e.g. DBSCAN (Ester *et al*., 1996)). The differences between monothetic vs. polythetic methods relate mainly to hierarchical clustering, where the former uses the inputs one by one and the latter all the inputs at once.

2.3. Comparisons of data and dimension reductions

When reviewing the literature on method comparisons, we first focus on dimension reduction methods and then on data reduction methods. The focus is on neutral evaluations of methods rather than evaluations in papers presenting novel methods. While papers presenting novel methods generally include an evaluation and conclude at least partial superiority of the invented method, such as some of those found in Section 2.2, they may be biased to a lesser or greater extent towards data and evaluation measures suitable for that particular method.

The large number of methods has obviously also stimulated a large number of performance comparisons between them. The comparisons mainly vary in terms of used data and evaluation measures, whereas there may be some variation in the precise utilization of methods. For instance, Flexer (1997, 2001) used Pearson correlation, Duch and Naud (1996) hypercubes in 3-5 dimensions and Bezdek and Pal (1995) the metric topology preserving index to show that MDS outperforms the SOM. Trosset (2008) argues that a serial combination of clustering and MDS is superior to the SOM. Venna and Kaski (2001) and Nikkilä *et al.* (2002) show superiority of the SOM and GTM in terms of trustworthiness of neighborhood relationships, while later Himberg (2004) and Venna and Kaski (2007) show superiority of CCA in terms of the same measure. Not surprisingly, de Vel *et al.* (1996) show, using Procruses analysis and Spearman rank correlation coefficients on various datasets, that the superiority of a method depends on the used evaluation measures and data. Lately, Lee and Verleysen (2009) proposed a unified measure based upon a co-ranking matrix for evaluating dimension reductions, a good starting point for generic evaluations. Lueks *et al.* (2011) further developed the measure by letting the user specify the properties that are more important to be preserved. While being useful aids in comparing methods, they neither show nor propose existence of one superior method for every type of data and preferences of similarity preservation. Hence, despite many attempts, inconsistent comparisons result in no conclusion on the superiority of one method.

When reviewing the literature on methods for data reduction, one can easily observe that neither there is a unanimity on the best available method. Herein, the focus is on comparisons between methods performing DDR and stand-alone data reduction methods. Bação *et al.* (2005) show that the SOM outperforms *k*-means clustering with 3 evaluation measures and 4 datasets. Flexer (1997; 2001) show that *k*-means clustering outperforms the SOM using a Rand index and 36 datasets. Waller *et al.* (1998) show on 2580 datasets that the SOM performs equally well with *k*-means clustering and better than other methods. Balakrishnan *et al.* (1994) show that *k*-means outperforms the SOM on 108 datasets, but do not decrease the SOM neighborhood to zero at the end of learning (as e.g. Kohonen (2001) proposes). Vesanto and Alhoniemi (2000) showed on 3 datasets that two-level clustering of the SOM is equally accurate as agglomerative and partitive methods, while obviously being computationally cheaper and having merits in visualizing relations in data. Ultsch and Vetter (1994) compare the SOM with hierarchical and *k*-means clustering and concludes that the SOM not only provides an equally accurate result, but also an easily interpretable output. Despite no unanimity on superiority, the literature still indicates that the SOM, and its adaptations, are equally considerable alternatives for data reduction as other methods, such as centroid-based and hierarchical clustering.

3. Methodology

The dimension reduction methods used in this paper are chosen to span different structure preservation objectives. In this section, we introduce three dimension reduction methods: metric MDS, Sammon's mapping and the SOM. The methods for data reduction are, on the other hand, chosen as to their resemblance to the clustering of the SOM and their suitability for data compression of dimension reductions. Again, we focus on three methods: VQ, *k*-means clustering and Ward's hierarchical clustering. VQ and *k*-means resembles the functioning of the prototype-based clustering of the SOM. Agglomeration of hierarchical clustering, while differing from the data compression of VQ, is suitable for DDR as it can be restricted by neighborhood relations of a dimension reduction and adapted to account for cluster size. This section sets a starting point for the comparative discussion by introducing the methods with an emphasis on their objective functions. In the sequel, if not otherwise mentioned, all distances are treated as Euclidean.

3.1. Dimension Reduction Methods

Dimension reductions employed in this paper have their basis in the first-generation methods. The main basis of classical methods is in the family of MDS-based projections. We first turn to a distance-preserving counterpart of PCA, classical metric MDS (Young and Householder, 1938; Torgerson, 1952). Due to its linearity constraints, we also turn to a non-linear MDS-based method, Sammon's (1969) mapping. That is, the aim is to project high-dimensional data x_j to a two-dimensional data vector y_j by preserving distances. Let the distance in the input space between x_j and x_h be denoted $d_x(j,h)$ and the distance in the output space between y_j and y_h be denoted $d_y(j,h)$. This gives us the objective function of metric MDS:

$$E_{MDS} = \sum_{k \neq l} \left(d_x(j,h) - d_y(j,h)\right)^2. \tag{1}$$

Sammon's mapping is an MDS method in that it also attempts to preserve pairwise distances between data but differs by focusing on local distances relative to larger ones. The square-error objective function for Sammon's mapping is

$$E_{SAM} = \frac{1}{\sum_{j \neq h} d_x(j,h)} \sum_{j \neq h} \frac{\left(d_x(j,h) - d_y(j,h)\right)^2}{d_x(j,h)} \tag{2}$$

and shows that it considers all pairs (j,h) normalized with the distance in the original space $d_x(j,h)$ and weighted with $1/d_x(j,h)$. The objective functions of MDS-based methods are optimized with an iterative steepest-descent process.

The SOM (Kohonen, 1982) differs by reducing both dimensions and data through a neighborhood-preserving VQ. In particular, the sequential SOM can be seen as a spatially constrained form of VQ. We follow, however, the batch SOM algorithm that can instead be seen as a spatially constrained counterpart of *k*-means clustering. The training algorithm proceeds according to two steps: (1) finding the best-matching units (BMUs) and (2) adjusting the prototype vectors m_i. To find $\min(d_x(j,i))$, the first step assigns the input data x_j (where j=1,2,…,N) to their BMU $m_b \in m_i$ (where i=1,2,…,M). Hence, data x_j are projected to an equidimensional prototype vector m_i, not a two-dimensional vector as in MDS. The second step adjusts each m_i with the batch updating algorithm:

$$m_i(t+1) = \frac{\sum_{j=1}^{N} h_{ib(j)}(t) x_j}{\sum_{j=1}^{N} h_{ib(j)}(t)}, \tag{3}$$

where t is a discrete time coordinate and the neighborhood $h_{ib(j)}$ often some decreasing function of neighborhood radii and time.

Mathematical treatment of the SOM has, however, shown to be difficult. Despite an extensive discussion of the form and existence of an objective function, the literature has still not provided one for the general case (Yin, 2008). It is, however, clear that a decomposed distortion measure illustrates the learning of the SOM (a discrete form with a fixed neighborhood of that suggested in Oja and Lampinen (1992)):

$$E_{SOM} = \sum_{j=1}^{N} \sum_{i=1}^{M} d_x(j,i)^2 h_{ib} d_y(i,b). \tag{4}$$

3.2. Data Reduction Methods

Data reduction and compression, or clustering, methods can be used for dividing data or prototypes into homogeneous groups. First, we introduce two clustering counterparts of the SOM. The sequential and batch SOMs can be seen as spatial counterparts of two prototype-based clustering methods: VQ (Linde *et al.*, 1980) and *k*-means clustering (MacQueen, 1967), respectively. VQ attempts to model the probability density functions in data x_j by prototype vectors m_i. As the SOM, it uses $\min(d_x(j,i))$ for finding the BMU for x_j, but then updates only the BMU towards the data vector. Hence, it attempts to minimize the standard squared error function, or quantization error:

$$J_{VQ} = \sum_{j=1}^{N} d_x(j,b)^2. \tag{5}$$

K-means is a similar least-square partitioning algorithm that pairs each data x_j to a cluster k (where k=1,2,...C) and then updates the centroids c_k to averages of all attracted data. That said, the aim is again to minimize the squared error function:

$$J_{km} = \sum_{j=1}^{N} \sum_{k=1}^{C} d_x(j,k)^2. \tag{6}$$

The third type of data reduction is hierarchical clustering. The following Ward's (1963) criterion is used as a basis for agglomerating clusters with the shortest distance:

$$d_{kl} = \frac{n_k n_l}{n_k + n_l} \cdot d_x(k,l)^2, \tag{7}$$

where k and l represent clusters, n_k and n_l the cardinality of clusters k and l, and $\left\| c_k - c_l \right\|^2$ the squared Euclidean distance between the cluster centers of clusters k and l. When clusters k and l are merged to cluster h, the cardinality n_h is the sum of n_k and n_l and the centroid c_h the mean of c_k and c_l weighted by n_k and n_l. Hence, this specification accounts for cluster size. A particularly advantageous feature of hierarchical methods is that agglomeration can be restricted to some specific property of the underlying similarity relations between clusters. For instance, the distance between non-adjacent clusters can be set to infinite, where adjacency needs to be defined. Again, clusters can be said to agglomerate as to minimize the Euclidean distance to the centroids, or the squared error function.

4. DDR Combinations for the Task at Hand

This section discusses specific aims, needs and restrictions of DDR combinations for visual financial performance analysis. From this discussion, we pinpoint dimensions of DDR combinations relevant for measuring suitability of methods for the task at hand.

4.1. Aims and Needs for the Task

The aim of models for visual financial performance analysis is to represent high-dimensional and big data of financial entities in easily understandable formats. Data and dimension reductions hold promise for the task, but the aim of the models still set some specific needs and restrictions. While recent advances in information technology have enabled access to databases with nearly endless amounts of financial information (e.g. Bankscope, Bloomberg, Standard & Poor's Capital IQ, etc), financial data are oftentimes problematic in being incomplete and non-normal (e.g. Deakin, 1976). For instance, in the case of representing a financial entity with its balance-sheet information, it is more common than not that some items of the balance sheet are missing. An example of skewed distributions is the commonly appearing power-law distribution or its variants. This derives to two necessities: the computational cost of the method needs to be considerably low and scalable and the method needs to be flexible for problematic data.

The main aim of the low-dimensional mappings is to use them as displays for additional information, in particular (a) individual data, (b) structural properties of data and (c) qualities of the models. This is due to three respective reasons: (a) the two-dimensional plane should function as a basis or display for visual performance comparisons of financial institutions (i.e. individual data); (b) for the human brain to recognize patterns in data, we need to provide guidance for interpreting general data structures, and oftentimes possess other linkable information as well; and (c) qualities of a dimension reduction may vary across mappings and locations in mappings as all information cannot be preserved in a lower dimension. The main aim of these mappings is hence not to be an end, but rather to function as a basis for a wide range of visualizations.

4.2. Aims and Needs of DDR Combinations

When evaluating or comparing performance of data and dimension reduction methods, particularly DDR combinations, quantitative measures have difficulties in accounting for qualitative differences in properties of methods. Hence, we translate in this section the needs for visual financial performance analysis into four qualitative dimensions for evaluating DDR combinations: form of structure preservation, computational cost, flexibility for problematic data and shape of the output.

Form of structure preservation As all relations in a high dimensional space can obviously not be preserved in a lower dimension, there are differences in what locations are stressed when preserving the structure. Given these differences, the main characteristics of structure preservation should obviously match important desires of the particular task at hand. The key question is thus: Which relations are of central importance for visual financial performance analysis? With a main focus on visualizing individual financial entities on a low-dimensional display, correctly locating neighboring financial institutions becomes particularly important. This leads to trustworthiness of neighborhood relationships being more important than precision on the precise distance to those far away. Noise and erroneous data as well as comparability issues related to reporting differences, for instance, also motivate attempting this type of a local order-preserving mapping rather than focusing on global detail.

Computational cost We oftentimes have access to extensive amounts of financial data in today's databases. This obviously sets some restrictions on computational cost and scalability of methods. While we acknowledge that computation time is not entirely a qualitative property, it has still not been incorporated in quantified evaluation measures. It is also worth to consider that computational expense is not only a one-off cost when creating a dimension reduction, but also when updating it. Combinations with data reduction methods may also affect the computational cost of a dimension reduction.

Flexibility for problematic data Methods differ in flexibility for non-normal and incomplete data, something more common than not in real-world financial settings. This said, desired properties of dimension reduction methods are flexibility for incomplete and non-normal data. While the former can be defined in terms of treatment of missing values, the latter depends largely on the task at hand. Most often data are pre-processed for ideal results, including treatment of skewed distributions. Yet, pre-processing seldom does, and is most often not desired to, compress the data into uniform density. Hence, one type of tolerance towards outliers can be derived from the output of methods.

Shape of the output One of the main aims is to use a dimension reduction as a display to which we link additional information. In particular, we use the low-dimensional mappings as displays for information, such as individual data and structural properties. This turns the focus to the shape of the outputs or mappings. Outputs of dimension reduction mappings can take a wide range of forms. We consider the interrelated properties of the shape to be: continuous vs. discrete mappings, optional vs. mandatory data reductions and pre-defined vs. data-driven grid shapes. While a mandatory data reduction is generally not desirable, we do not consider it a significant disadvantage due to the large amounts of available data. This leads also to restricting mappings to discrete rather than continuous, whereas continuous mappings would obviously be desirable. The largest difference for interpretation, especially in terms of linking visualizations, is between pre-defined and data-driven grid shapes. While methods with data-driven grid shapes may better adapt to data, the methods with pre-defined grid shapes are superior in functioning as a regularly formed display for additional information.

5. Qualitative and Quantitative Experiments

This section introduces data and dimension reductions to financial performance analysis. We first discuss the relative goodness of methods for the task from a qualitative perspective. Then, we illustrate how data and dimension reduction methods can be utilized for visual financial performance analysis of European banks.

For dimension reduction, we consider the SOM, metric MDS and Sammon's mapping. Metric MDS and Sammon's mapping belong to the group of distance-preserving methods, while the SOM belongs to the group of topology-preserving methods with a pre-defined grid shape. For data reduction, we consider Ward's hierarchical clustering, *k*-means clustering and VQ. Due to access to overabundant amounts of financial data, we concentrate herein mainly on serial and parallel combinations of these stand-alone methods for DDR combinations. While the SOM performs DDR by default, large SOMs can be further enhanced by a second-level clustering of its prototypes. Conceptually, the functioning of the SOM differs from the other DDR combinations as the two tasks of data and dimension reduction are treated as concurrent subtasks. In serial combinations, the dimension reduction is always subordinate to the data reduction, while parallel combinations deal separately with the initial dataset.

5.1. A Qualitative Comparison

This section pinpoints the qualitative discussion of DDR combinations for visual performance analysis to four key issues: form of structure preservation, computational cost, flexibility for problematic data and shape of the output.

Form of structure preservation The main difference between DDR combinations is how the dimension reduction methods differ in the properties of data they attempt to preserve. For our task, this leads to one key question: Which methods better assure trustworthy neighbors? MDS-based methods with objective functions attempting distance preservation, while likely being better at approximating distance structures, may end up with skewed errors across the projection. To this end, it has been shown that the SOM, which stresses neighborhood relations, better assures trustworthy neighbors (Venna and Kaski, 2001; Nikkilä *et al.*, 2002). That is, data found close-by each other on a SOM display are more likely to be similar in terms of the original data space as well. The conceptual difference in structure preservation between distance- and topology-preserving methods is illustratively described by Kaski (2001) with an experiment on a curved two-dimensional surface in a three-dimensional space: the former methods may follow the surface in data with two dimensions, whereas the latter require three dimensions to describe the structure.

Computational cost Expensive computations is obviously an issue when dealing with large financial datasets. Generally, computing pairwise distances between data is costly and the order of magnitude is N^2. The topology preservation of the SOM relates instead to the grid size M with an order of magnitude of M^2 (Kaski, 1997). This implies that if M=N then the methods have similar complexity, but more importantly that grid size M can be adjusted for cheaper complexity. Further, parallel DDR combinations suffer from an additional computational cost as the clustering is performed on the initial dataset rather than on a reduced number of prototypes. The computational cost of MDS-based methods motivates serial DDR combinations. Another issue related to computational cost is the lack of an explicit mapping function for the MDS-based methods. Hence, when including new samples the projection needs to be recomputed. While new samples can be visualized via projection to their best-matching data, each update requires recomputing the projection.[2] In contrast, the SOM can be cheaply updated with individual data using a sequential algorithm.

Flexibility for problematic data The methods significantly differ in flexibility for problematic data. Methods dealing with distance preservation have obvious difficulties with incomplete data. However, the SOM, and its self-organization, can be seen as tolerant to missing values by only considering the available ones in matching (Samad and Harp, 1992). In practice, the SOM has been shown to be robust when up to approximately 1/3 of the variables are missing (Kohonen, 2001; Kaski and Kohonen, 2001; Denny and Squire, 2005, Sarlin, 2011). Indeed, the SOM has even been shown to be effective for imputing missing values (e.g. Cottrell and Letrémy, 2005). Tolerance towards outliers is measured in terms of representation of skewed distributions. An

[2] While relative MDS (Naud and Duch, 2000) allows to add new data to the basis of an old MDS, it does still not update all distances within the mapping.

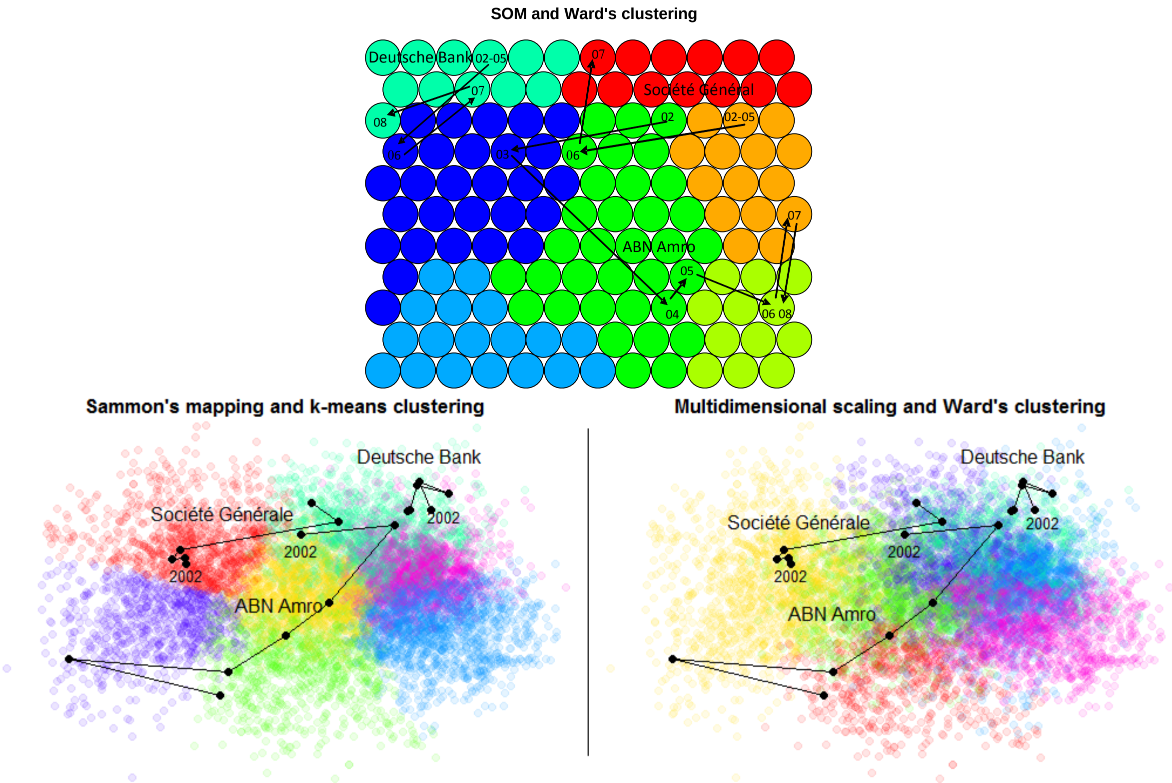
MDS-based mapping becomes difficult to interpret if it is stretched towards directions of outliers. While the processing of the SOM does not *per se* treat outliers, its regularly shaped grid of prototypes facilitates visualizing data with non-uniform density functions.

Shape of the output A key to using a dimension reduction as a display, and linking information to it, is the shape of its output. While the SOM has a discrete mapping, mandatory data reduction and predefined grid shape, MDS-based methods are its contrasts by having continuous mappings, optional data reduction and data-driven lattice (if combined with data reduction). The predefined SOM grid, while also having drawbacks for representing structural properties of data, facilitates the interpretation of linked information. Today, it is standard that the SOM comes with a wide set of linked extensions for visual analytics, such as the so-called component planes, U-matrix and frequency plots (Vesanto, 1999). Even though visual aids for showing distance structure and density compensate for constraints set by the grid shape, there is a large group of other aids that enhance the representation of available information in data. The visual aids, while not always being even applicable, have generally not been explored in the context of MDS-based projections. Component planes, for instance, are difficult to visualize due to the lack of a reduced number of prototypes. Even DDR combinations with serial VQ, i.e. processing similar to that of the SOM, would still lack the concept of neighborhood relations of a regularly shaped grid.

5.2. Illustrative Experiments

The qualitative discussion of properties of DDR combinations for financial performance analysis still lacks illustrative examples. Here, we show experiments with the above discussed methods. Dimension reduction is performed with the SOM, metric MDS and Sammon's mapping and data reduction with Ward's hierarchical clustering, *k*-means clustering and VQ. We explore various combinations for DDR for achieving easily interpretable models for visual financial performance analysis. The methods are chosen and combined as to their suitability for data reduction of dimension reductions, and *vice versa*.

Data The dataset used in this paper consists of annual financial ratios for banks from the European Union, including all provided financial ratios in the Bankscope database. The pre-processing of this dataset follows the suggested framework in Eklund *et al*. (2011). The dataset consisted initially of 38 financial ratios for 1,236 banks spanning from 1992:12–2008:12. We chose to use 24 ratios by dropping those with more than 25% missing data. Observations with more than 1/3 of missing values were removed. Finally, we were left with a resulting 9,655 rows of data, and a total of 855 banks. Although the SOM is tolerant to missing values, we need to impute missing data as distance-preserving methods require complete data. We use the SOM for imputing missing values. A SOM allows mapping incomplete data to their BMUs by only considering the available variables. Hence, complete data were used for training a SOM, incomplete data were mapped to their BMUs and the missing values were imputed from their BMUs. Moreover, although outliers are not a problem *per se*, they still affect the interpretability of the models. Not to lose significant amounts of data, we use modified boxplots for trimming with replacement. The modified boxplot is preferred over Winsorizing, for instance, as it accounts for variable-specific distributions, resulting in replacement of a total of 7.39% of the data, distributed as needed per variable and tail.



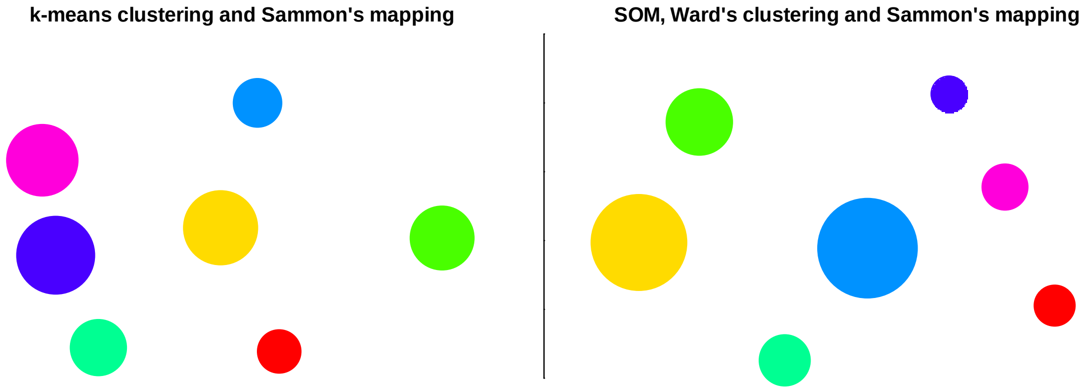
Notes: The figures show parallel DDR combinations on the entire dataset; Sammon's mapping is combined with k-means clustering, and MDS and the SOM are combined with Ward's clustering.. Color codes on each mapping correspond to clusters and the superimposed trajectories to a performance comparison of three large European banks from 2002-2008.

Figure 2. Parallel DDR combinations applied to the financial dataset, as well as trajectories for three large European banks from 2002–08.

Parallel DDR Figure 2 shows parallel DDR combinations on the entire dataset. Sammon's mapping is combined with *k*-means clustering, and MDS and the SOM are combined with Ward's clustering.[3] Ward's clustering of the SOM is, however, performed on its prototypes rather than on the dataset and restricted to agglomerate only adjacent clusters in the SOM topology. We do not, however, consider this option for MDS-based projections as there is no natural definition of adjacency. On top of all three mappings, we superimpose cluster color coding and a performance comparison of trajectories from 2002-2008 for three large European banks. The projections of MDS and Sammon's mapping on this large dataset are very similar, whereas *k*-means clustering has less overlapping cluster memberships in the mapping than Ward's clustering. The trajectories as well as the underlying variables confirm that, while the orientations of the two MDS-based projections are somewhat different from those of the SOM

[3] When training SOMs, one has to set a number of parameters. We mainly follow the decision-framework developed in Eklund *et al.* (2011). We use a set of quality measures to track the topographic and quantization accuracy as well as clustering of the map. The map format is chosen to be 75:100, as it approximates the recommended ratio of the two largest eigenvalues (see e.g. Kohonen, 2001).

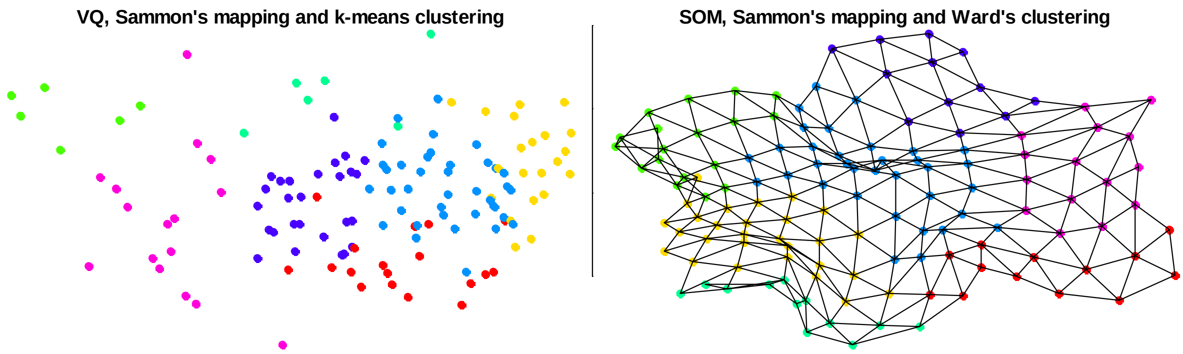
model, their structure is still very similar. Yet, the computational cost differs significantly. While it takes on an ordinary personal computer only a few seconds to train SOM-based models on these data, the MDS-based projections require several hours on a dedicated server.



Notes: The figure shows serial DDR combinations on the entire dataset; Sammon's mapping is combined with k-means clustering, and the SOM with second-level Ward's clustering.. Color codes on each mapping correspond to clusters. Not to clutter the display, trajectories are not displayed in this figure.

Figure 3. Serial DDR combinations applied to the financial dataset.

Serial DDR For cheaper complexity, we further explore possibilities of DDR by testing serial combinations. Figure 3 shows a Sammon's mapping of the *k*-means cluster centroids as well of the second-level centroids of the SOM, where size represents the number of data in each cluster. This type of usage of MDS-based methods was already proposed by Sammon (1969) due to their high computational cost, and later applied by Flexer (2001), for instance. It is, indeed, a cheap way to illustrate relations between the cluster centroids, but lacks detail for structural as well as individual analysis.



Notes: The figures show serial and parallel DDR combinations on the entire dataset; Sammon's mapping is combined with VQ and k-means clustering, and the SOM with Ward's clustering.. Color codes on each mapping correspond to clusters and the net-like representation illustrates neighborhood relations.

Figure 4. Serial and parallel DDR combinations applied to the financial dataset.

Serial and parallel DDR Costly but detailed MDS-based projections in Figure 2 and cheap but very crude projections in Figure 3 motivate finding a compromise solution. For reducing computational expense, we still end up relying on a serial DDR combination. For more detail, however, we attempt to reduce the initial dataset to a smaller but representative dataset. This type of data compression can be achieved with standard VQ that approximates probability density functions of data. The compressed prototypes can then be used as an input for a parallel DDR. Conceptually, while still lacking the interaction between the tasks as well as the regular grid shape, we come close to what is achieved using a SOM in Figure 2 by relying on both serial and parallel DDR combinations. The left plot in Figure 4 shows a VQ of the initial dataset and then a subsequent Sammon's mapping and *k*-means clustering on the VQ prototypes. The right plot in Figure 4 shows a corresponding Sammon's mapping of SOM prototypes with a superimposed cluster color coding. However, the figure illustrates two issues: the ordered SOM prototypes have less overlap of cluster memberships and the importance of naturally defined topological relations. The former issue is most likely a result of interaction between the tasks of data and dimension reduction, but might also benefit from the inclusion of neighborhood relations when agglomerating clusters. The latter issue of a regularly shaped grid is particularly useful when attempting to visualize as much of the available information as possible through linked visualizations.

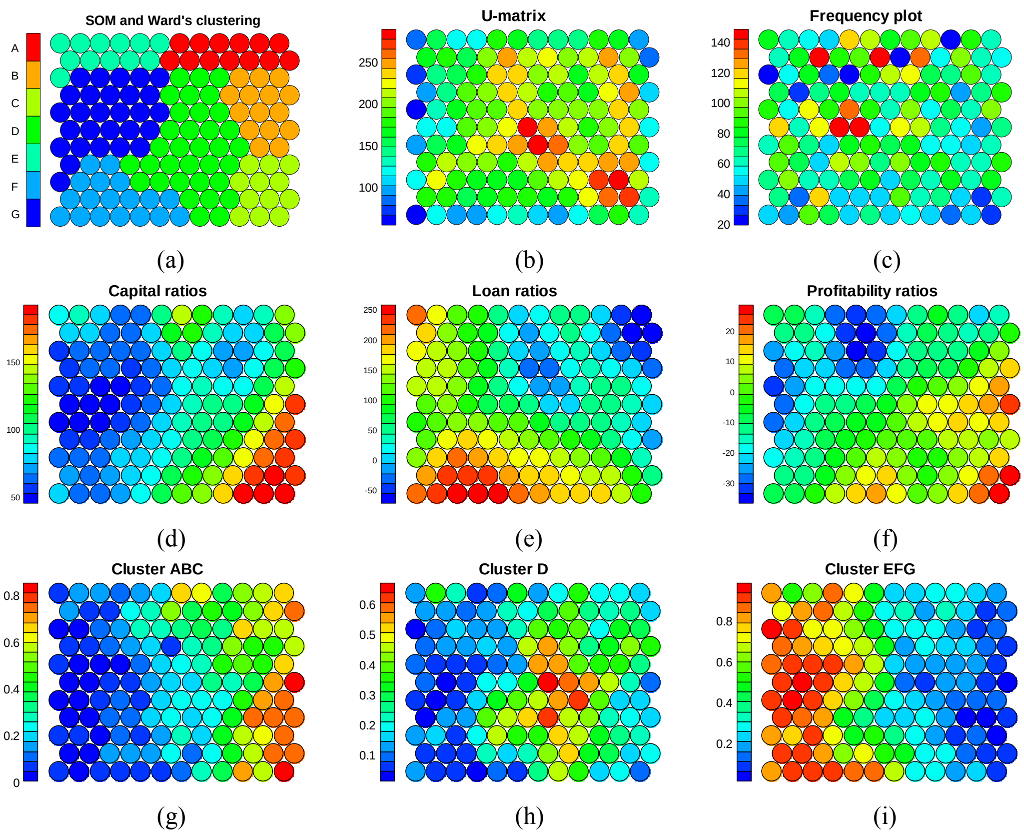
5.3. The SOM and its visualization aids

This section first briefly reviews visualization aids for the SOM and then illustrates the use of the regularly shaped SOM grid, and its visualization aids. Figure 2 showed the two-dimensional SOM grid, and trajectories for three large European banks from 2002–08, but a central question remains: How should we interpret the map? The possibility of linking additional information to the SOM grid has stimulated the development of a wide scope of visualization aids (see Vesanto (1999) for an early overview). Generally, the aids fall into one of three groups: (1) those compensating structural properties inherent in data that the regular grid shape eliminates, (2) those extending the visualization of properties inherent in data but not normally accessible in dimension reductions, and (3) those linking the SOM grid with other methods or data to enhance the understanding of the task.

The first group includes means to represent the distance structure and density on a SOM, something missing due to the VQ and grid shape. Densities on the SOM are generally assessed with frequency plots and the Pareto density estimation matrix (P-matrix) (Ultsch, 2003a). Examples of aids for assessing distance structures are Sammon's mapping, the Unified distance matrix (U-matrix) (Ultsch and Siemon, 1990) and cluster connections (Merkl and Rauber, 1997). Moreover, some methods attempt to account for both structures and densities, such as the U*-matrix (Ultsch, 2003b), the sky metaphor visualization (Latif and Mayer, 2007), the neighborhood graph (Poelzlbauer *et al.*, 2005) and smoothed data histograms (Pampalk *et al.*, 2002).

The second group consists of visualizations that enhance the representation of multidimensional information. Component planes are a standard method for visualizing the spread of values of individual dimensions on the SOM, but they have been further enhanced in several aspects. For instance, Vesanto and Ahola (1999) use a SOM for reorganizing the component planes according to correlations and Neumayer *et al.* (2007) introduced the metro map discretization to summarize all component planes onto one plane. Kaski *et al.* (2001) have developed a visualization of the contribution of each variable to distances between prototypes, that is, the cluster structure. Another extension, while partly also belonging to the other groups, is visualization of vector fields (Poelzlbauer *et al*., 2006) for assessing contributions to the cluster structure and for finding correlations and dependencies in the underlying data.

The third group uses other methods or data for further enhancing the understanding of the task. One common way to represent cluster structures in a SOM is applying a second-level clustering on the prototypes, and visualizing it through color coding (Vesanto and Alhoniemi, 2001). This has also been done using fuzzy methods (Sarlin and Eklund, 2011), where soft memberships of prototypes in each cluster can be represented on their own grids or associated to individual data. The strengths and directions of state transition patterns on a SOM can similarly be represented on own grids, such as prototype-to-cluster transition probabilities in Sarlin *et al*. (2012), not to mention using the prototypes as an input for some other predictive method, such as the neural network in Serrano Cinca (1996), and visualizing the prediction on the SOM grid. An example of combining the display of a SOM with other data is to link network relations between entities on the grid (Sarlin, 2012). In a financial setting, this enables relating the state of financial performance of a cross section of entities to their network relations, i.e. systemic relevance.



Notes: The figures link additional information to the regularly shaped SOM grid. Charts (a-c) illustrate structural properties of the model: (a) shows cluster memberships of the second-level clustering, (b) shows average distances between prototypes, or the so-called U-matrix, and (c) shows the frequency distribution on the SOM grid. Charts (d-f) show the spread of three subdimensions of financial performance on the SOM grid: capital, loan and profitability ratios. Charts (g-i) show for each prototype the probability of transition to three areas on the SOM grid: Clusters A, B and C, cluster D, and clusters E, F and G.

Figure 5. An exemplification of linking information to the SOM grid.

Here, we show some examples of how visualizations from the above three groups can be linked to the SOM. The previously presented Figure 2 already showed a financial performance comparison over time of three large European banks using labels and trajectories. Figure 5 uses the regular shape of the SOM grid as a basis for nine different representations of additional information. First, Figures 5a, 5b and 5c illustrate structural properties of the model: (a) shows hard cluster memberships of the second-level clustering, (b) shows distance structures using a u-matrix visualization, and (c) shows the frequency distribution on the SOM grid. While Figures 5a and 5b show similar characteristics of cluster structures, Figure 5c shows that the upper-left corner of the map is comparatively more dense. Second, Figures 5d, 5e and 5f enable assessing correlations and distributions by showing the spread of three financial performance measures on the SOM grid: capital, loan and profitability ratios. Here, one can observe that, generally, the right part represents well-performing and the left part poor banks. Third, Figures 5g, 5h and 5i illustrate temporal transition patterns by showing for each prototype the probability of transition to three areas on the SOM grid: clusters A, B, and C (good); cluster D (average); and clusters E, F and G (poor banks). The figures show that, while the left part (poor) is more stable than the right (good), the center is a transition cluster. Hence, the first group is represented by Figures 5b and 5c, the second group by Figures 5d, 5e and 5f, and the third group by 5a, 5g, 5h and 5i.

6. Discussion

This paper has considered data and dimension reduction methods, as well as their combination, for visual financial performance analysis. The discussions and illustrations in Section 5, while being at times somewhat trivial, are motivated by inconsistency of argumentation for and application of various methods. The main conclusion of the comparison is that the SOM has several useful properties for financial performance analysis. In particular, we have noted the following advantages of the SOM over alternative distance-preserving methods: 1) trustworthy neighbors, 2) low computational cost, 3) flexibility for problematic data, and 4) a regularly shaped grid. It is, however, worth noting that the relative goodness of a method depends always on the task in question. That said, the SOM is obviously far from a panacea for all sorts of data and dimension reduction. When only attempting stand-alone tasks, it is indeed very likely that there exists better methods than the SOM. Similarly, when attempting DDR, the superiority of one method over others depends entirely on the aims of the task in question.

Even though the SOM has been assessed as superior for visual financial performance analysis, it is worth to carefully consider its limitations:

1. The SOM performs a crude mapping. Rather than data points, the SOM attempts to embed the prototypes. A significant constraint if detail is of central importance.

2. The regular grid shape sets some restrictions on the SOM. For instance, it may cause interpolating sparse locations with idle prototypes, lead to an analyst overinterpreting the regular-like y and x axes, and lead to the need for additional visual aids to fully represent structures.

3. Mathematical treatment of the SOM has shown to be problematic. The lack of an objective function, as well as a general training schedule for or proof of convergence, complicates assessing convergence of a SOM.

The comparison in this paper has covered classical first-generation dimension reduction methods. The main question that remains is: Can the results of this comparison be generalized to all available methods? As reviewed in Section 2, CCA has been shown to outperform the SOM in terms of trustworthiness of neighborhood relations (Himberg, 2004; Kaski and Venna, 2007). However, it falls short in the shape of the output. It is thus important to consider methods from the second generation with the key properties of the SOM. There are two conceptually similar topology-preserving methods that possess DDR capabilities and a pre-defined grid shape: GTM and XOM. GTM mainly differs from the SOM by relying on well-founded statistical properties based upon Bayesian learning with an objective function, namely the log-likelihood, which is optimized by the Expectation-maximization algorithm. This objective function directly facilitates assessing convergence of the GTM. Even though Bishop *et al.* (1998b) originally stated that the GTM is computationally comparable to the SOM, it has later been shown that the SOM is cheaper (e.g. Rauber *et al.*, 2000). This may result from the number of developed algorithmic shortcuts for computing SOMs, such as fast-winner search (Kaski, 1999). Both methods are flexible for problematic data, i.e. outliers and missing values, through a similar pre-defined grid shape and an extension of the GTM for treating missing values (Carreira-Perpiñan, 2000; Sun *et al.*, 2001). However, while choosing parameters for the SOM may be a tedious task, given adequate initializations and parameterizations, convergence has seldom appeared to be a problem in practice (Yin, 2008). A decade after the introduction of the GTM, neither it nor its variants, such as the S-Map (Kiviluoto and Oja, 1998), have displaced the standard SOM.

The XOM is a computational framework for data and dimension reduction. By inverting the functioning of the SOM, the XOM systematically exchanges functional and structural components of topology-preserving mappings by self-organized model adaptation to the input data. It has two main advantages compared to the SOM: (a) reduced computational cost, and (b) applicability to non-metric data as there is no restriction on the distance measures. Even though the use of non-metric dissimilarity measures is of little use on the data in this paper, while still having potential for other pairwise financial data, the reduced computational cost is particularly beneficial for large financial datasets in general. The XOM has, however, been recently introduced and is thus still lacking thorough tests in relation to other methods, such as comparisons to SOMs with algorithmic shortcuts. Yet, the XOM should be considered as a valid alternative to the SOM paradigm.

It is worth to note, as widely suggested (e.g. Trosset, 2008; Lee and Verleysen, 2007), that one of the main reasons for the SOM being very popular for a broad range of tasks, such as classification, clustering, visualization, prediction, missing value imputation etc, might be because it produces an intuitive output using a simple and an intuitive principle. This simplicity, while being beneficial for a method to be widely accepted, applied and understood, should still not be used for assessing relative goodness. One should, nevertheless, note that when introducing dimension reductions to the general public, such as policy- or decision-makers in general, simplicity is definitely an asset. To this end, this paper argues that the optimal method for financial performance analysis is one from the family of methods that perform a topology-preserving mapping to a regularly shaped and predefined grid, out of which a method has to be chosen according to one's own preferences.

7. Conclusions

Every task has its own needs. Data and dimension reduction for financial performance analysis should thus be performed with methods that have the best overall suitability for that task, not the best data-processing capabilities for some other objective. To this end, this paper has addressed the choice of method for visual financial performance analysis from a qualitative perspective. We have first discussed the properties of three inherently different classical first-generation dimension reduction methods, and their combination with data reduction, and illustrated their performance in a real-world financial application. The conclusions drawn from the comparison of classical methods was then prolonged to second-generation methods. The qualitative discussion and experiments showed superiority of the SOM for financial performance analysis in terms of four dimensions: form of structure preservation, computational cost, flexibility for problematic data and shape of the output. When considering second-generation methods, the recently introduced the GTM and the XOM have clear potential for similar tasks. GTM improves the SOM paradigm with its well-defined objective function, but is computationally more costly, while XOM is a recently introduced promising method, but still lacks thorough comparisons.

From the discussions in this paper, an obvious conclusion is that the family of methods that perform a topology-preserving mapping to a regularly shaped and predefined grid provides the means for visual financial performance analysis. The aims and needs for the task at hand, where the main focus lies on using the output as a display for additional information in general and individual data in particular, are neither rare objectives in other fields. While not being generalizable to their full extent, parts of the conclusions herein will also apply in other fields and domains. The methods advocated in this paper do not, however, provide a panacea for visual financial performance analysis. They should be paired with other methods, not least visualizations of different kinds, that compensate for missing properties when having, for instance, a regularly shaped grid. It also motivates exploring the information commonly linked to the SOM in not only the same family of methods with predefined grid shapes, but also other dimension reduction paradigms in general.

Acknowledgements

The author wants to thank Barbro Back and Tomas Eklund, and seminar participants at the Data Mining and Knowledge Management Laboratory on 13 February 2012 and IS Seminar at Åbo Akademi University on 1 March 2012 for useful comments.

References

Aigner, W., Miksch, S., Möller, W., Schumann, H., Tominski, C., 2008. Visual Methods for Analyzing Time-Oriented Data. IEEE Transactions on Visualization and Computer Graphics 14(1), pp. 47-60.

Bação, F., Sousa Lobo, V., Painho, M., 2005. Self-organizing Maps as Substitutes for K-Means Clustering. In Proceedings of the International Conference on Computational Science : 476-483

Back, B., Sere, K., Vanharanta, H., 1998. Managing Complexity in Large Data Bases using Self-Organizing Maps. Accounting Management and Information Technologies 8(4), pp. 191-210.

Balakrishnan, P.V., Martha, C.C., Varghese S.J., Phillip A.L., 1994. A Study of the Classification Capabilities of Neural Networks Using Unsupervised Learning: A Comparison with K-Means Clustering. Psychometrika, 59, pp. 509-525.

Belkin, M., Niyogi, P., 2001. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds.) Advances in Neural Information Processing Systems 14, pp. 586–691, MIT Press, Cambridge, MA.

Bishop, C.M., Svensson, M., Williams, C.K.I., 1997. GTM: A principled alternative to the self-organising map. In: M. C. Mozer, M. I. Jordan, and T. Petsche (Eds.) Advances in Neural Information Processing Systems 9, MIT Press, Cambridge, MA.

Bishop, C.M., Svensson, M., Williams, C.K.I., 1998a. GTM: the Generative Topographic Mapping. Neural Computation 10(1), pp. 215-234.

Bishop, C.M., Svensson, M., Williams, C.K.I., 1998b. Developments of the generative topographic mapping. Neurocomputing 21(1–3), pp. 203-224.

Carreira-Perpiñan, M.A., 2000. Reconstruction of sequential data with probabilistic models and continuity constraints. In: S.A. Solla, T.K. Leen, K.-R. Müller (Eds.) Advances in Neural Information Processing Systems 12, pp. 414-420, MIT Press, Cambridge, MA.

Cottrell, M., Letrémy, P., 2005. Missing values: Processing with the Kohonen algorithm. In: Proceedings of Applied Stochastic Models and Data Analysis (ASMDA 2005), Brest, pp. 489–496.

Deakin, E.B., 1976. Distributions of Financial Accounting Ratios: Some Empirical Evidence. The Accounting Review 51, pp. 90-96.

Demartines, P., Hérault, J., 1997. Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. IEEE Transactions on Neural Networks 8, pp. 148-154.

Demirgüç-Kunt, A., Detragiache, E., 2000. Monitoring Banking Sector Fragility. A Multivariate Logit. World Bank Economic Review 14(2), pp. 287-307.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society (Series B) 39(1), pp. 1-38.

Demyanyk, Y.S., Hasan, I., 2010. Financial crises and bank failures: a review of prediction methods. Omega 38(5), pp. 315-324.

Duch, W., Naud, A., 1996. Multidimensional scaling and Kohonen's self-organizing maps. In: Proceedings of the 2nd Conference on Neural Networks and their Applications, Szczyrk, Poland, pp. 138-143.

Eklund, T., Back, B., Vanharanta, H., Visa, A., 2003. Using the Self-Organizing Map as a Visualization Tool in Financial Benchmarking. Information Visualization 2(3), pp. 171-181.

Eklund, T., Sarlin, P., Jokipii, A., 2011. Visual Financial Benchmarking using the Self-Organizing Map: A Revisit. Turku Centre for Computer Science Technical Report, No. 1009.

Ezzamel, M., Mar-Molinero, C., 1987. On the distributional properties of financial ratios, Journal of Business Finance and Accounting 14, pp. 463–481.

Ester, M., Kriegel, H-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: E. Simoudis, J. Han, U.M. Fayyad (Eds.) Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226-231.

Flexer, A., 1997. Limitations of Self-Organizing Maps for Vector Quantization and Multidimensional Scaling. In: M.C. Mozer (Eds.) Advances in Neural Information Processing Systems 9, pp.445-451, MIT Press, Cambridge, MA.

Flexer, A., 2001. On the Use of Self-organizing Maps for Clustering and Visualization. Intelligent Data Analysis 5(5), pp. 373-384.

Fraser, D.R., Phillips, W., Rose, P.S., 1974. A Canonical Analysis of Bank Performance. The Journal of Financial and Quantitative Analysis 9(2), pp. 287-295.

González-Hermosillo, B., 1999. Determinants of Ex-Ante Banking System Distress: A Macro-Micro Empirical Exploration of Some Recent Episodes. IMF Working Paper No. 99/33.

Halkos, G.E., Salamouris, D.S., 2004. Efficiency measurement of the Greek commercial banks with the use of financial ratios: a data envelopment analysis approach. Management Accounting Research 15, pp. 201-224.

Himberg, J., 2004. From insights to innovations: data mining, visualization, and user interfaces. PhD thesis, Helsinki University of Technology, Espoo, Finland.

Kaski, S., 1997. *Data exploration using self-organizing maps*. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82: Espoo.

Kaski, S., 1999. Fast winner search for SOM based monitoring and retrieval of high dimensional data. In: Proceedings of the IEEE International Conference on Artificial Neural Networks (ICANN'99), London, UK, pp. 940-945.

Kiviluoto K., Oja E., 1998. S-map: A network with a simple self-organization algorithm for generative topographic mappings. In: M.I. Jordan, M.J. Kearns, S.A. Solla (Eds.) Advances in Neural Information Processing Systems 10, pp. 549-555, MIT Press, Cambridge, MA.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biological Cybernetics 66, pp. 59–69.

Kohonen, T., 2001. *Self-Organizing Maps*, 3rd edition. Springer-Verlag, Berlin.

Kruskal, J.B.,1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, pp. 1-27.

Kung, C.Y., Kun-Li W., 2007. Applying Grey Relational Analysis and Grey Decision-Making to evaluate the relationship between company attributes and its financial performance—A case study of venture capital enterprises in Taiwan. Decision Support Systems 43(3), pp. 842-852.

Lampinen, J., Oja, E., 1992. Clustering Properties of Hierarchical Self-Organizing Maps. Journal of Mathematical Imaging and Vision 2(2-3), pp. 261–272.

Latif, K., Mayer, R., 2007. Sky-Metaphor Visualisation for Self-Organising Maps. In: Proceedings of the 7th International Conference on Knowledge Management (I-KNOW 2007), Graz, Austria.

Linde, Y., Buzo, A., Gray, R.M., 1980. An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications 28(1), pp. 702--710.

Lee, J.A., Verleysen, M., 2007. *Nonlinear dimensionality reduction*. Springer, Information Science and Statistics series.

Lee, J.A., Verleysen, M., 2009. Quality assessment of dimensionality reduction: Rank-based criteria. Neurocomputing 72(7-9), pp. 1431-1443.

Lueks, W., Mokbel, B., Biehl, M., Hammer, B., 2011. How to Evaluate Dimensionality Reduction?. In: Proceedings of the Workshop on New Challenges in Neural Computation, Frankfurt, Germany.

van der Maaten, L.J.P., Hinton, G.E., 2008. Visualizing high-dimensional data using t-SNE. Journal of Machine Learning Research 9, pp. 2579–2605.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297. Berkeley, CA: University of California Press.

Mar-Molinero, C., Serrano-Cinca, C., 2001. Bank failure: a multidimensional scaling approach. European Journal of Finance 7(2), pp. 165-183.

Martin-del-Brio, B., Serrano-Cinca, C., 1993. Self-organizing neural networks for the analysis and representation of data: Some financial cases. Neural Computing & Applications 1(3), pp. 193-206

Merkl, D., Rauber, A., 1997. Alternative Ways for Cluster Visualization in Self-Organizing Maps. In: Proceedings of the Workshop on Self-Organizing Maps (WSOM 1997), Helsinki, Finland.

Männasoo, K., Mayes, D.G., 2009. Explaining bank distress in Eastern European transition economies. Journal of Banking & Finance 33(2), pp. 244-253.

Naud, A., Duch, W., 2000. Interactive data exploration using MDS mapping. In: Proceedings of the 5th Conference on Neural Networks and Soft Computing, Zakopane, Poland, pp. 255-260.

Neumayer, N., Mayer, R., Poelzlbauer, G., Rauber, A., 2007. The metro visualisation of component planes for self-organising maps. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN 2007), Orlando, FL, USA, IEEE Computer Society.

Nikkilä, J., Törönen, P., Kaski, S., Venna, J., Castrén, E., Wong, G., 2002. Analysis and visualization of gene expression data using self-organizing maps. Neural Networks 15(8-9), pp. 953-966.

Nobes, C., 2006. The survival of international differences under IFRS: towards a research agenda. Accounting and Business Research 36(3), pp. 233–245.

Pampalk, E., Rauber, A., Merkl, D., 2002. Using Smoothed Data Histograms for Cluster Visualization in Self-Organizing Maps. In: Proceedings of the International Conference on Artificial Neural Networks (ICANN 2002), pp.871-876, Madrid, Spain.

Pearson, K., 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine 2(6), pp. 559-572.

Poelzlbauer, G., Dittenbach, M., Rauber, A., 2006. Advanced visualization of Self-Organizing Maps with vector fields. Neural Networks 19(6–7), pp. 911-922.

Poelzlbauer, G., Rauber, A., Dittenbach, M., 2005. Advanced visualization techniques for self-organizing maps with graph-based methods. In: Proceedings of the Second International Symposium on Neural Networks (ISNN 2005), pp 75-80, Chongqing, China, Springer Verlag.

Rauber, A., Paralic, J., Pampalk, E., 2000. Empirical evaluation of clustering algorithms. Journal of Information and Organizational Sciences 24 (2), pp. 195-209.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, pp. 2323-2326.

Samad, T., Harp, S.A., 1992. Self-organization with partial data. Network: Computation in Neural Systems 3, pp. 205–212.

Sammon, J.W., 1969. A Non-Linear Mapping for Data Structure Analysis. IEEE Transactions on Computers 18(5), pp. 401–409.

Sarlin P, 2012. Chance Discovery with Self-Organizing Maps: Discovering Imbalances in Financial Networks. In: Y. Ohsawa and A. Abe (Eds.) Advances in Chance Discovery, Berlin, Springer Verlag, pp. 49-61.

Sarlin, P., Eklund, T., 2011. Financial Performance Analysis of European Banks using a Fuzzified Self-Organizing Map. In: Proceedings of the 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES'11), Kaiserslautern, Germany, September 12--14, pp. 183-192

Sarlin, P., Peltonen, T.A., 2011. Mapping the State of Financial Stability. ECB Working Paper 1982.

Sarlin, P., Yao, Z., Eklund, T., 2012. Probabilistic Modeling of State Transitions on the Self-Organizing Map: Some Temporal Financial Applications. In: Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS 2012), pp. 1128--1137, Hawaii, USA.

Seçme, N., Bayrakdaroğlu, A., Kahraman, C., 2009. Fuzzy performance evaluation in Turkish Banking Sector using Analytic Hierarchy Process and TOPSIS. Expert Systems with Applications 36(9), pp. 11699-11709.

Serrano-Cinca, C., 1996. Self Organizing Neural Networks for Financial Diagnosis. Decision Support Systems 17, pp. 227-238.

Shepard, R.N., 1962. The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika, 27, pp. 125-140; 219-246.

Sun, Y., Tino, P., Nabney, I.T., 2001. GTM-based data visualisation with incomplete data. Technical Report. Neural Computing Research Group, Birmingham, UK.

Tenenbaum, J.B., de Silva, V., Langford, J. C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science, 290, pp. 2319-2323.

Torgerson, W. S., 1952. Multidimensional scaling: I. Theory and method. Psychometrika 17, pp. 401-419.

Trosset, M.W., 2008. Representing Clusters: K-Means Clustering, Self-Organizing Maps, and Multidimensional Scaling. Technical Report 08-03, Department of Statistics, Indiana University.

Ultsch, A., 2003a).Maps for the Visualization of high-dimensional Data Spaces. In: Proceedings of the Workshop on Self-Organizing Maps (WSOM 2003), Kyushu, Japan, pp. 225–230.

Ultsch, A., 2003b. U*-Matrix: a Tool to visualize Clusters in high dimensional Data. Technical Report No. 36, Deptartment of Mathematics and Computer Science, University of Marburg, Germany.

Ultsch, A., Siemon, H.P., 1990. Kohonen's self organizing feature maps for exploratory data analysis. In*:* Proceedings of the International Conference on Neural Networks (ICNN 1990), Dordrecht, the Netherlands, pp. 305–308.

Ultsch A., Vetter, C., 1994. Self-Organizing-Feature-Maps versus Statistical Clustering Methods: A Benchmark, University of Marburg, FG Neuroinformatik & Kuenstliche Intelligenz, Research Report 0994.

Venna, J., Kaski, S., 2001. Neighborhood preservation in nonlinear projection methods. An experimental study. In: Proceedings of the International Conference on Artificial Neural Networks (ICANN'01) pp. 485-491. Berlin, 2001. Springer.

Venna, J., Kaski, S., 2007. Comparison of visualization methods for an atlas of gene expression data sets. Information Visualization 6(2), pp. 139-154.

Vesanto, J., 1999. SOM-Based Data Visualization Methods. Intelligent Data Analysis 3(2), pp. 111-126.

Vesanto, J., Ahola, J., 1999. Hunting for Correlations in Data Using the Self-Organizing Map. In: Proceeding of the International ICSC Congress on Computational Intelligence Methods and Applications (CIMA'99), ICSC Academic Press, pp. 279-285.

Vesanto J., Alhoniemi, E., 2000. Clustering of the Self-Organizing Map. IEEE Transactions on Neural Networks 11(3), pp. 586-600.

Waller, N.G., Kaiser, H.A., Illian, J.B., Manry, M., 1998. A comparison of the classification capabilities of the 1-dimensional Kohonen Neural Network with two partitioning and three hierarchical cluster analysis algorithms. Psychometrika 63, pp. 5-22.

Ward, J., 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58, pp. 236–244.

De Vel, O., Lee, S., Coomans, D., 1996. Comparative Performance Analysis of Non-Linear Dimensionality Reduction Methods. In: D. Fischer, H-J. Lenz (Eds.) Learning From Data: Artificial Intelligence and Statistics V, New York, Springer, pp. 320-345.

Weinberger, K.Q., Saul, L.K., 2005. Unsupervised learning of image manifolds by semidefinite programming. International Journal of Computer Vision 70(1), pp. 77-90.

Wismüller, A., 2009. A computational framework for non-linear dimensionality reduction and clustering. In: J. Principe, R. Miikkulainen (Eds.) Advances in Self-Organizing Maps, Lecture Notes in Computer Science, vol. 5629, Springer, 2009, pp. 334–343.

Yin, H., 2008. The Self-Organizing Maps: Background, Theories, Extensions and Applications. In: J. Fulcher, L. Jain (Eds.) Computational Intelligence: A Compendium, Springer Berlin/Heidelberg, pp. 715-762.

Young, G., Householder, A. S., 1938. Discussion of a set of points in terms of their mutual distances. Psychometrika 3, pp. 19-22.

Zopounidis, C., Despotis, D.K., Stavropoulou, E., 1995. Multiattribute evaluation of greek banking performance. Applied Stochastic Models and Data Analysis 11, pp. 97-107.

Turku Centre *for* Computer Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi

University of Turku
- Department of Information Technology
- Department of Mathematics

Åbo Akademi University
- Department of Information Technologies

Turku School of Economics
- Institute of Information Systems Sciences

ISBN 978-952-12-2754-7
ISSN 1239-1891