Visual Terrain Analysis of High-Dimensional Datasets

Wenyuan Li¹, Kok-Leong Ong², and Wee-Keong Ng¹

¹ Nanyang Technological University, Centre for Advanced Information Systems,
Nanyang Avenue, N4-B3C-14, Singapore 639798
liwy@pmail.ntu.edu.sg, awkng@ntu.edu.sg

² School of Information Technology, Deakin University,
Waurin Ponds, Victoria 3217, Australia
leong@deakin.edu.au

Abstract. Most real-world datasets are, to a certain degree, skewed. When considered that they are also large, they become the pinnacle challenge in data analysis. More importantly, we cannot ignore such datasets as they arise frequently in a wide variety of applications. Regardless of the analytic, it is often that the effectiveness of analysis can be improved if the characteristic of the dataset is known in advance. In this paper, we propose a novel technique to preprocess such datasets to obtain this insight. Our work is inspired by the resonance phenomenon, where similar objects resonate to a given response function. The key analytic result of our work is the *data terrain*, which shows properties of the dataset to enable effective and efficient analysis. We demonstrated our work in the context of various real-world problems. In doing so, we establish it as the tool for preprocessing data before applying computationally expensive algorithms.

1 Introduction

The subfield of data analysis is essentially a collection of algorithms that focused on analyzing large datasets of high-dimensionality. Often than not, the cornerstone of these algorithms is to address the dimensionality curse when trying to provide effective and efficient results for a given user query. Towards this, there have been many research done; including cluster analysis to find clusters embedded in subspaces (also known as *subspace clustering* or *biclustering*), and dimensionality reduction.

In cluster analysis, most models are based on distance or similarity measures, or correlation measures of feature subsets or objects. While they unveil the details of subspace clusters, most are of no interest to the user. For example, more than 10,000 clusters were obtained through OP-clustering [1] on a drug activity dataset with a dimension of $10,000 \times 30$. Clearly, this is overwhelming to the user trying to find insights about the data in question, e.g., the relationship among patterns rather than a list of patterns. Usually, closer inspection would suggest close relationships among clusters. And if high level insights is what the user is after, then this level of pattern redundancy would be inappropriate. Yet, a combinatorial explosion of patterns (satisfying the query) occur as the size and dimensionality of the dataset increases. Dimension reduction is one alternative to ‘curb’ the combinatorial explosion of patterns by passing a reduced space to the analytical algorithms. The drawback, however, is the loss of patterns embedded

Fig. 1. The rating matrix for 4 reviewers and 5 movies, biclusters and PCA 2-dimensional space

in the subspace of the original space. This happens because most reduction techniques made use of distance or similarity measures over the full dimension, and therefore lack the mechanism to find the embedded patterns that are subtle but important.

In this paper, we introduce the concept of *data terrain* to visualize high-dimensional datasets while overcoming the limitations of subspace clustering and dimension reduction. Our proposal effectively reveals the relationship among subspace clusters, and allows the user to explore the data at different levels of details. We show, by means of real-world applications (e.g., biclusters, outliers, and frequent itemsets), how the data terrain can help discover generic patterns that can be utilized to effectively analyze the patterns embedded in the original space. Unfortunately, to find this data terrain under varying conditions proved to be NP-hard. Thus, our contribution in this paper includes the proposal of efficient techniques to find the data terrain. We next show a motivating example to illustrate the relevance of data terrains in analysis. We then introduce the resonance model in Section 3 and summarize our work in Section 4.

2 Motivating Example

We begin by introducing the concept of data terrain and show by means of an example, how it facilitates better data analysis; and why it is better than other techniques like biclustering and dimension reduction. Our example is based on the survey of popular movies. Fig. 1(a) shows the rating matrix W of 4 reviewers (R_i) on 5 movies (M_j), where each movie is rated on a scale of 1 to 20.

We first use biclustering to analyze the relationship between the reviewers and the movies. If requiring biclusters with at least 2 rows and columns, more than 10 biclusters can be discovered. Fig. 1(c) – (f) are the distinct biclusters found in this case. While these biclusters precisely characterized the reviewers’ ‘rating style’ on movies, there is too much redundancy in the solution for such a small dataset. In real-world situations where the dataset is much larger, it will take much longer before the an-

	M_2	M_5	M_3	M_4	M_1
R_1	20	16	15	8	5
R_4	12	20	20	5	3
R_2	8	2	5	19	16
R_3	6	3	7	17	18

Plain ← → Mountain

(a) matrix W' of reordered W

(b) 3D of W'

(c) bicluster 1

(d) bicluster 2

(e) bicluster 3

(f) bicluster 4

(g) BLOCKS and actual data

(h) Approximation of the actual terrain

Fig. 2. (a) – (f): Reordered rating matrix given by Fig. 1; (g) & (h): Illustrations of σ -BLOCKs

alyst is able to come to a conclusion. Worse, the analyst is likely to become confused on the real relationship between the reviewers and the movies if they were to work at this level of detail. The other approach would be to use dimension reduction techniques. We consider using a popular technique known as Principal Component Analysis (PCA). The PCA's output is shown in Fig. 1(g) and (h). Again, the relationship between the reviewers and movies is revealed. However, as PCA runs across the full dimension of the data, we lose many subtle and local insights about reviewers and movies. For example, movies M_2 , M_4 and M_5 are perceived as different in Fig. 1(g). Yet, if we check back on our analysis using biclustering, we can see from Fig. 1(d) that they are actually quite similar if we consider the ratings from reviewer R_2 and R_3 . Thus, if only PCA is performed, we will not be able to arrive at this conclusion.

Interestingly, if we view W in 3D space, we can capture the relationships that both biclustering and PCA revealed. As Fig. 1(b) shows, a direct 3D 'plot' of W does not seem to reveal any interesting insights – but if we were to reorder W into W' as shown in Fig. 2(a) (and also Fig. 2(c) – (f); where every bicluster can be shown in this manner [1]), we have a 3D terrain of W' as depicted in Fig. 2(b). Notably, this terrain provides the insights that earlier requires both biclustering and PCA analysis.

To illustrate this, notice that any bicluster from Fig. 2(c) – (f) can be obtained by selecting some points from the 'mountains' and 'plains' in this terrain. At the same time, we can also make conclusions that would otherwise be obtained through PCA: (i) there are primarily two groups of movies and reviewers; (ii) M_3 and M_5 have higher similarity than M_2 despite being in the same group; and (iii) the 'rating style' of R_2 and R_3 is opposite that of R_1 and R_4 . Thus, the terrain captures both local and global relationships about the data in an intuitive and effective manner. Of course, real-world datasets are much more complex that result in more complicated terrains. Consequently, trying to discover such a terrain proved to be a NP-hard problem.

3 Discovering Data Terrains

Conceptually, moving from W to W' is simply the reordering of the matrix to form the ‘mountains’ and ‘plains’. Yet, this ordering on both dimensions can be difficult to achieve efficiently on massive datasets. To prove the hardness of this problem, we first give the following definitions. Let $\mathcal{O}$ be a set of objects, where $o \in \mathcal{O}$ is defined by a set of attributes $\mathcal{A}$. Further, let w_{ij} be the magnitude of o_i over $a_j \in \mathcal{A}$. Then we can represent the relationship of all objects and their attributes in a matrix $W = (w_{ij})_{|\mathcal{O}| \times |\mathcal{A}|}$ for the weighted bipartite graph $G = (\mathcal{O}, \mathcal{A}, E, W)$, where E is the set of edges. Thus, discovering the ‘mountains’ transforms into the problem of evaluating subgraphs where the magnitude of all its edges are above some ‘altitude’, i.e., $w_{ij} \geq \sigma$. Formally, the concept of a ‘mountain’ in this data terrain is called a BLOCK.

Definition 1. *Given a weighted bipartite graph G , a σ -BLOCK (or simply σ -B) is a subgraph $G' = (\mathcal{O}', \mathcal{A}', E', W')$ of G satisfying $w_{ij} \geq \sigma$ for any $i \in \mathcal{O}'$ and $j \in \mathcal{A}'$.*

From Definition 1, σ -B can be intuitively viewed as a plane (or a transverse section) with a specified altitude σ that ‘cuts’ across W . In the case of Fig. 1(b), we set the plane at $\sigma=10$ to obtain two 10-Bs as shown in Fig. 2(b): $\{R_1, R_4, R_2\} \times \{M_2, M_5\}$ and $\{R_2, R_3\} \times \{M_4, M_1\}$. Therefore, a series of σ -Bs can be generated when considering planes with different σ values. Once this set of BLOCKs relevant to G is found, we can order them to find the data terrain.

Definition 2. *Given a bipartite graph $G = (\mathcal{O}, \mathcal{A}, E, W)$ and a set of BLOCKs $\{B_1, B_2, \dots, B_k\}$ found from G , the terrain of W is two ordered sequences of $\mathcal{O}$ and $\mathcal{A}$, such that these BLOCKs are placed consecutively in the reordered W .*

It is interesting to note that sorting both dimensions, i.e., $\mathcal{O}$ and $\mathcal{A}$, is an extension of sorting a single dimensional array to determine its distribution. However, sorting both dimensions simultaneously to get the 2-dimensional distribution is practically infeasible, i.e., finding the σ -BLOCKs by iteratively decreasing σ from the maximum value of W is NP-hard. In fact, finding a single σ -B is NP-hard.

Theorem 1. *Finding the largest σ -BLOCK ($|\mathcal{O}'| \times |\mathcal{A}'|$) is NP-hard.*

Proof. Our problem can be reduced from the *maximum edge biclique* [2], which is NP-complete. Details of this proof can be referred to [3].

Given the difficulty of finding σ -Bs, we seek alternative methods to discover the data terrain. Since our objective is to find the ‘mountains’ and ‘plains’ but *not* where they are on the terrain, then some approximation to the actual terrain (that is computationally efficient) should suffice. The insignificance of the specific locations of the ‘mountains’ and ‘plains’ can be demonstrated from Fig. 2(g) and (h), where the same set of insights are obtained from both figures. As this terrain is approximated, we called it the macro-view¹. To obtain the macro-view of a terrain for a dataset, we used a novel

¹ The complete work of this paper includes a *micro-view* of the data terrain. Together, they provide a complete solution for analysis of high-dimensional datasets. Due to space constraints, the reader is referred to [3] for the details.

Fig. 3. The model and the effectiveness of the linear instance

model inspired by the physics of resonance. This resonance model is very efficient even on very large and high-dimensional datasets. Instead of checking every σ -B, we can simulate a resonance experiment by injecting a response function to elicit objects of interest to the analyst. Proofs of all theorems in this section are omitted and refer to [3].

3.1 The Model

To simulate a resonance phenomenon, we require a forcing object $\tilde{o}$, such that when an appropriate response function $\mathbf{r}$ is applied, $\tilde{o}$ will resonate to elicit those objects $\{o_i, \dots\} \subset \mathcal{O}$ in G , whose ‘natural frequency’ is similar to $\tilde{o}$. This ‘natural frequency’ represents the characteristics of both $\tilde{o}$ and the objects $\{o_i, \dots\}$ who resonated with $\tilde{o}$ when $\mathbf{r}$ was applied. For the weighted bipartite graph $G = (\mathcal{O}, \mathcal{A}, E, W)$ and $W = (w_{ij})_{|\mathcal{O}| \times |\mathcal{A}|}$, this ‘natural frequency’ of $o_i \in \mathcal{O}$ is $\mathbf{o}_i = (w_{i1}, w_{i2}, \dots, w_{i|\mathcal{A}|})$. Since a one-dimensional array (or vector) can be sorted to obtain its own terrain, we also refer $\mathbf{o}_i$ as the terrain of the object o_i . Likewise, the terrain of the forcing object $\tilde{o}$ is defined as $\tilde{\mathbf{o}}_i = (\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_{|\mathcal{A}|})$.

Put simply, if two objects of the same ‘natural frequency’ will resonate and therefore, should have a similar terrain. The evaluation of resonance strength between objects o_i and o_j is given by the response function $\mathbf{r}(\mathbf{o}_i, \mathbf{o}_j) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$. We defined this function abstractly to support different measures of resonance strength. For example, one existing measure to compare two terrains is the well-known *rearrangement inequality theorem*, where $\mathbf{I}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i$ is maximized when the two positive sequences $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ are ordered in the same way (i.e. $x_1 \geq x_2 \geq \dots \geq x_n$ and $y_1 \geq y_2 \geq \dots \geq y_n$) and is minimized when they are ordered in the opposite way (i.e. $x_1 \geq x_2 \geq \dots \geq x_n$ and $y_1 \leq y_2 \leq \dots \leq y_n$).

Notice if two vectors maximizing $\mathbf{I}(\mathbf{x}, \mathbf{y})$ are put together to form $M = [\mathbf{x}; \mathbf{y}]$ (in MATLAB format), we obtain the terrain. More importantly, all σ -Bs are immediately obtained from this terrain with the need to search every σ -B! This is why the model is efficient – it only needs to consider the resonance strength among objects once the appropriate response function is selected. For example, the response function $\mathbf{I}$ is a suitable candidate to characterize the similarity of terrains of two objects. Likewise, $\mathbf{E}(\mathbf{x}, \mathbf{y}) = \exp(\sum_{i=1}^n x_i y_i)$ is also an effective response function.

To find the ‘mountains’ and ‘plains’, the forcing object $\tilde{o}$ evaluates the resonance strength of every objects o_i against itself to locate a ‘best fit’ based on the contour of its terrain. By running this iteratively, those objects that resonated with $\tilde{o}$ are discovered

and placed together to form the ‘mountains’ within the 2-dimensional matrix W . In the same fashion, the ‘plains’ are discovered by combining those objects that resonated weakly with $\tilde{o}$. This iterative learning process between $\tilde{o}$ and G is outlined below.

Initialization. Set up $\tilde{o}$ with a uniform distribution: $\tilde{o} = (1, 1, \dots, 1)$; normalize it as $\tilde{o} = \text{norm}(\tilde{o})^2$; then let $k = 0$; and record this as $\tilde{o}^{(0)} = \tilde{o}$.

Apply Response Function. For each object $o_i \in \mathcal{O}$, compute the resonance strength $r(\tilde{o}, o_i)$; store the results in a vector $\mathbf{r} = (r(\tilde{o}, o_1), r(\tilde{o}, o_2), \dots, r(\tilde{o}, o_{|\mathcal{O}|}))$; and then normalize it, i.e., $\mathbf{r} = \text{norm}(\mathbf{r})$.

Adjust Forcing Object. Using $\mathbf{r}$ from the previous step, adjust the terrain of $\tilde{o}$ for all $o_i \in \mathcal{O}$. To do this, we define the adjustment function $c(\mathbf{r}, \mathbf{a}_j) : \mathbb{R}^{|\mathcal{O}|} \times \mathbb{R}^{|\mathcal{O}|} \rightarrow \mathbb{R}$, where the weights of the j -th attribute is given in $\mathbf{a}_j = (w_{1j}, w_{2j}, \dots, w_{|\mathcal{O}|j})$. For each attribute a_j , $\tilde{w}_j = c(\mathbf{r}, \mathbf{a}_j)$ integrates the weights from $\mathbf{a}_j$ into $\tilde{o}$ by evaluating the resonance strength recorded in $\mathbf{r}$. Again, c is abstract, and can be materialized using the inner product $c(\mathbf{r}, \mathbf{a}_j) = \mathbf{r} \bullet \mathbf{a}_j = \sum_i w_{ij} \cdot r(\tilde{o}, o_i)$. Finally, we compute $\tilde{o} = \text{norm}(\tilde{o})$ and record it as $\tilde{o}^{(k+1)} = \tilde{o}$. We denote the resonance model as $R(\mathcal{O}, \mathcal{A}, W, \mathbf{r}, c)$, where the instances of functions $\mathbf{r}$ and c can be either $\mathbf{I}$ or $\mathbf{E}$.

Test Convergence. Compare $\tilde{o}^{(k+1)}$ against $\tilde{o}^{(k)}$. If the result converges, go to the next step; else apply $\mathbf{r}$ on $\mathcal{O}$ again (i.e., forcing resonance), and then adjust $\tilde{o}$.

Macro-View of Terrain. Sort the objects $o_i \in \mathcal{O}$ by the coordinates of $\mathbf{r}$ in descending order; and sort the attributes $a_i \in \mathcal{A}$ by the coordinates of $\tilde{o}$ in descending order.

3.2 Properties of the Model

The abstract view of the general model is given in Fig. 3(a). Depending on the response and adjustment function, the abstract model instantiates into different implementations. In practice, we have the linear model $R(\mathcal{O}, \mathcal{A}, W, \mathbf{I}, \mathbf{I})$, and the non-linear model $R(\mathcal{O}, \mathcal{A}, W, \mathbf{E}, \mathbf{E})$. We shall discuss some important properties of our model in this section. In particular, we show that the model gives a good approximation to the actual terrain, and that its iterative process converges quickly.

Approximation to Actual Terrain. Using the synthetic data from Fig. 2(g), we can see how well both implementations approximate the actual terrain. The linear and non-linear model converges to a precision of $\epsilon = 0.001$, i.e., once $\|\tilde{o}^{k+1} - \tilde{o}^k\| \leq \epsilon$, terminates. The reordered matrices are the same as Fig. 2(h). We then performed the same test on the movie-rating example. Result of linear model is shown in Fig. 3(b) and the non-linear model in Fig. 2(a). Obviously, $R(\mathcal{O}, \mathcal{A}, W, \mathbf{E}, \mathbf{E})$ gives a better approximation, where the ‘mountains’ and ‘plains’ are easily distinguishable. Thus, we can conclude that the different instances of R may give an approximate of the actual terrain. These conclusions are also empirically proven in [3].

Convergence. Since the resonance model is iterative, it is essential that it converges quickly to be efficient. Essentially, the model can be seen as a type of *discrete dynamical system* [4]. The convergence of linear and non-linear models is proven below.

Theorem 2. $R(\mathcal{O}, \mathcal{A}, W, \mathbf{r}, c)$, where $\mathbf{r}, c$ are $\mathbf{I}$ or $\mathbf{E}$, converges in limited iterations.

² $\text{norm}(\mathbf{x}) = \mathbf{x} / \|\mathbf{x}\|_2$, where $\|\mathbf{x}\|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$ is 2-norm of vector $\mathbf{x} = (x_1, \dots, x_n)$.

Fig. 4. A case of data analysis by macro-view of the terrain: visualization of MovieLens data (with 943 users, 1,682 movies and 100,000 ratings on the scale of 1 to 5. It is available at <http://www.grouplens.org>), and analysis of its local terrain. We obtain the macro-view using $R(\mathcal{O}, \mathcal{A}, W, \mathbf{I}, \mathbf{I})$ to get the terrain in (a). It is possible that the user is not satisfied with an overview of the dataset and might be interested in further analysis. A case of ‘zoom in’ on the ‘crowded’ but a local terrain of the macro-view is shown in (b).

In practice, the model is very efficient because we are only interested in the convergence of orders of coordinates in $\tilde{\mathbf{o}}^k$ and $\mathbf{r}^k$. With k iterations, the complexity is $\mathcal{O}(k \times |\mathcal{O}| \times |\mathcal{A}|)$. In our experiments, our model converges within 50 iterations even on the non-linear configurations giving a time complexity of $\mathcal{O}(|\mathcal{O}| \times |\mathcal{A}|)$. In all cases, the complexity is sufficiently low to efficiently handle large datasets.

Average Inter-resonance Strength $\frac{1}{\binom{k}{2}} \sum_{\substack{i \neq j \in \mathcal{O}' \\ |\mathcal{O}'|=k}} \mathbf{r}(\mathbf{o}_i, \mathbf{o}_j)$ **among Objects.** Theorem 3 is in fact an optimization process to find the best k objects, whose average inter-resonance strength is the largest among any subset of k objects. Next, we exploit this property to unveil the relationship between the macro-view of the data terrain and the biclusters.

Theorem 3. *Given the macro-view terrain W' , the average inter-resonance strength $\frac{1}{\binom{k}{2}} \sum_{1 \leq i \neq j \leq k} \mathbf{r}(\mathbf{o}_i, \mathbf{o}_j)$ of the first k objects, w.r.t. the resonance strength with $\tilde{\mathbf{o}}$, is largest for any subset with k objects.*

Approximation to Maximum Edge Biclique (MEB). The non-linear configuration of our model, i.e., $R(\mathcal{O}, \mathcal{A}, W, \mathbf{E}, \mathbf{E})$ has such capability. Details refer to [3].

3.3 Real World Examples

A demonstration of how a macro-view of the data terrain can help the user in analysis is shown by a real-world case in Fig. 4. Next we show how it can have applications in data mining for finding frequent itemsets and biclustering in theory. All empirical evidences refer to [3].

Finding Frequent Itemsets. A transaction dataset can be constructed as a matrix, where each transaction is an object, and each item is an attribute whose value w_{ij} in $W_{|\mathcal{O}| \times |\mathcal{A}|}$ is 1 if the j -th item occurs in the i -th record, and 0 otherwise. We therefore have the following that relates frequent itemsets and BLOCKs.

Theorem 4 (Frequent Itemsets and BLOCKs). *A frequent itemset is the attribute set of a 1-BLOCK, and its support is the number of objects in the BLOCK.*

Discovering Biclusters. A popular measure for biclusters [5] is defined as Eqn. (1). The residue $H(W)$ of given a matrix W is a δ -bicluster if $H(W) \leq \delta$.

$$H(W) = \frac{1}{mn} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} (w_{ij} - w_{iJ} - w_{IJ} + |W|)^2 \quad (1)$$

where $w_{iJ} = \frac{1}{n} \sum_{j=1}^n w_{ij}$, $w_{IJ} = \frac{1}{m} \sum_{i=1}^m w_{ij}$, and $|W| = \frac{1}{mn} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} w_{ij}$.

Theorem 5 (Bicluster and Average Resonance Strength of Macro-View Terrains). *Given a matrix $W = (w_{ij})_{m \times n}$, where $\mathcal{O}$ are the rows and $\mathcal{A}$ are columns, we have the inverse relation of the average inter-resonance strength and $H(W)$ as follows*

$$H(W) = \|W\|^2 + |W|^2 - \frac{1}{n} \bar{\mathbf{r}}(W) - \frac{1}{m} \bar{\mathbf{r}}(W^T) \quad (2)$$

where $\|W\| = \sqrt{\frac{1}{mn} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} w_{ij}^2}$, $\bar{\mathbf{r}}(W) = \frac{1}{\binom{m}{2}} \sum_{\substack{1 \leq i, j \leq m \\ i \neq j}} \mathbf{I}(\mathbf{w}_{i-}, \mathbf{w}_{j-})$ is the average inter-resonance strength among $\mathbf{w}_{i-}$, and $\bar{\mathbf{r}}(W^T) = \frac{1}{\binom{n}{2}} \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \mathbf{I}(\mathbf{w}_{-i}, \mathbf{w}_{-j})$ is the average inter-resonance strength among $\mathbf{w}_{-j}$, and $\mathbf{w}_{i-}$ is the i -row vector of W with $\mathbf{w}_{-j}$ the j -column vector of W .

It can be interpreted as follows. Since $\|W\|$ and $|W|$ are sum of W in different forms, we can consider them as fixed constant. If the average inter-resonance strength of W and W^T , i.e., $\bar{\mathbf{r}}(W)$ and $\bar{\mathbf{r}}(W^T)$, is higher, then $H(W)$ is lower and thus, W behaves like a bicluster. For $R(\mathcal{O}, \mathcal{A}, W, \mathbf{I}, \mathbf{I})$ and $R(\mathcal{A}, \mathcal{O}, W^T, \mathbf{I}, \mathbf{I})$, we conclude that if we select the first k rows and columns of W with large resonance strength $\mathbf{r}(\mathbf{o}_i, \tilde{\mathbf{o}})$ to form W' , it is straightforward that we will have a smaller $H(W')$ and thus, W a bicluster.

4 Summary

In this paper, we proposed the data terrain as a means to visualize and analyze high-dimensional datasets. With this terrain, patterns in subspaces can be visualized and analyzed. We provided a novel solution to obtain the the macro-view of a terrain efficiently, and demonstrated its real-world application.

References

1. Liu, J., Wang, W.: Op-cluster: Clustering by tendency in high dimensional space. In: Proceedings of ICDM, Melbourne, Florida (2003) 187
2. Peeters, R.: The maximum edge biclique problem is NP-complete. *Disc. App. Math.* **131** (2003) 651–654
3. Li, W., Ong, K.L., Ng, W.K.: Visual terrain analysis of high dimensional datasets. Technical Report (www.deakin.edu.au/leong/tr0406) (TRC04/06), Deakin University (2005)
4. Sandefur, J.T.: *Discrete Dynamical Systems*. Oxford: Clarendon Press (1990)
5. Cheng, Y., Church, G.M.: Biclustering of expression data. In: Proceedings of the 8th International Conference on Intelligent System for Molecular Biology. (2000)