# Evaluating Entropic Based Clustering Algorithms on Biomedical Data

Jorge M. Santos
*INEB, Biomedical Engineering Institute*
*ISEP, School of Engineering, Polytechnic of Porto*
*Porto, Portugal*
*jms@isep.ipp.pt*

Frederico Morais
*ISEP, School of Engineering, Polytechnic of Porto*
*Porto, Portugal*

***Abstract*—Clustering algorithms are being widely used on biomedical data. They aim to extract important information that can be used to improve life conditions by helping specialized technicians on the decision process.**

**Clustering algorithms based on information theory concepts claim that by using higher order statistic they are able to extract more information from the data and therefore provide much better results.**

**In this work we try to verify this claim by comparing the performance of some entropic clustering algorithms against more conventional ones. Results of the performed experiments are not conclusive but they seem to indicate that this kind of entropic algorithms may provide some improvements when clustering biomedical data.**

## I. Introduction

In the past decades, a huge amount of research has been conducted using medical or biomedical data. Researchers have tried to extract information from this particular kind of data using methods of machine learning or computational intelligence. One of the research areas consists on trying to categorize, segment or distinguish different groups on a given dataset. This approach usually known as clustering helps on the decision process that in biomedical field assumes special importance.

One of the main problems in the clustering process arrives when data possesses some unusual structure causing the discovery of different groups much more difficult than usual. Basic clustering algorithms like k-means usually deal easily with globular or hiper-ellipsoidal groups. If the structure of the data is more complex it is expected that these kind of algorithms present worst results.

Since Shannon's work on information theory [1] that researchers try to apply entropy and related concepts on different disciplines and particularly to machine learning. Several entropic based clustering algorithms were also created and developers claim that their algorithms achieve much better results in different kinds of datasets. In this work we try to evaluate the usefulness of entropic clustering algorithms on biomedical data. We compare them against three spectral clustering algorithms, one density based clustering algorithm and also against k-means.

This paper is organized as follows: in the next section we make a brief overview on clustering; in Section III we describe the clustering algorithms used in this comparison; in Section IV we present the datasets and the metric used to evaluate the clustering performance and in the final sections we present the results, the discussion and we draw some conclusions.

## II. Brief overview on clustering

Clustering is an unsupervised learning process of finding structure in data. More specifically, it aims to group objects based on their similarity and to discriminate groups based on its elements dissimilarity. Clustering is consider to be a NP problem with no unique solution what makes clustering tasks difficult to implement and to evaluate. As opposite to classification problems where metrics based on class labels may be used to evaluate classification results, data usually used in clustering is not tagged and results must be evaluated using some clustering validation metrics [2]. There are three approaches to investigate cluster validity: external, internal and relative criteria. To avoid the use of these criteria one can evaluate the performance of clustering algorithms by using labeled data usually used on classification. By doing this we can compare the clustering results with the class labels using some specific performance indexes such as the Advanced Rand Index that we will describe later.

Clustering algorithms can usually be divided in several groups based on their clustering method: partitional, hierarchical, density-based and grid-based algorithms. Hierarchical clustering algorithms can further be divided into agglomerative or divisive algorithms according respectively to the decreasing or increasing number of clusters they produce at each step. K-means is an example of a partitional algorithm.

Entropy and related concepts are being used in clustering following different strategies. Entropy can be used for instance as a measure of intra or inter-cluster evaluation, as an objective function combined or not with other clustering methods or as a measure to compute similarity or dissimilarity matrices. One of the main disadvantages of entropic clustering algorithms is their higher computational complexity when compared with more traditional ones.

Clustering can be used in different areas like Image Processing and Pattern Recognition, Market segmentation,

Document classification, Gene analysis or Sociometry data analysis.

## III. Clustering algorithms used in the experiments

### A. Entropic Clustering Algorithms

Entropic clustering algorithms make use of information theoretical concepts such as mutual information (MI), Shannon's entropy, Renyi's entropy or even Havrda-Charvat's entropy to perform clustering. Entropic clustering tries to use the higher order statistics to extract higher order information from the data. Examples of entropic based clustering algorithms are the COOLCAT [3], that performs clustering on categorical data, the MECA algorithm [4], that combines entropy with fuzzy clustering, the minimum entropic clustering (MEC) [5] that tries to minimize entropy and uses an initial partition given by another clustering algorithm, the LEGCLust [6], that combines hierarchical and graph approaches, or even the more recent minCEntropy (MCE) [7] that uses the conditional entropy for quantifying both clustering quality and distinctiveness.

In our experiments we make use of LEGClust, minimum entropic clustering (MEC) and minCEntropy (MCE). In the following we present each one of these three algorithms.

**LEGClust** is an hierarchical algorithm that uses graphs based on an entropic dissimilarity matrix. It computes Rényi's quadratic entropy between pairs of points to evaluate the dissimilarity between them and build an entropic dissimilarity matrix. Based on this matrix an hierarchical algorithm is used to build the clusters using an aglomerative approach. Rényi's quadratic entropy is estimated by

$$\hat{H}_{R_2} = -\log\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(x_i - x_j; 0, 2h^2 I)\right) \quad (1)$$

where $N$ is the number of points $x$, $G(.)$, a Gaussian kernel, $m$, the dimension of $x$, $h$, the Parzen window size and $I$ the identity matrix. LEGClust algorithm builds a dissimilarity matrix by computing the entropy of every point in a given neighborhood. In a second stage it builds a proximity matrix constituted by layers of sub-graphs based on the dissimilarity matrix. In the final stage and using a hierarchical approach it creates clusters using the proximity matrix. One of the drawbacks of this algorithm is that it needs to tune 3 parameters: the neighborhood size, the Parzen window size and the minimum number of connections to build the clusters. One of the advantages of LEGClust is that it can produce arbitrarily shaped clusters.

**Minimum Entropic Clustering (MEC)** is an algorithm that uses Havrda-Charvat's conditional entropy and the following objective function:

$$\hat{J} = \begin{cases} 1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} p^{\alpha}(c_j|x_i) & \alpha > 1 \\ -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} p^{\alpha}(c_j|x_i)\ln(c_j|x_i) & \alpha = 1 \\ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} p^{\alpha}(c_j|x_i) - 1 & 0 < \alpha < 1 \end{cases} \quad (2)$$

where $n$ is the number of points, $m$ is the number of clusters and $p(c_j|x_i)$ is the conditional probability that is estimated by means of the Parzen window method and also by k-nearest neighbors. This is an iterative algorithm that tries to reduce conditional entropy starting by a initial partition given by k-means. In each step of the algorithm a rearrangement of the data is performed by shifting a point from a cluster to another and computing the conditional entropy. If there is a reduction the point is assigned to the new cluster. The algorithm stops until there is no rearrangement that improves entropy. Three parameters are needed: the number of clusters for the initial partition, the Parzen window size and the $\alpha$ value that determines which formula to use in (2).

**minCEntropy (MCE)** is a partitional clustering algorithm with an objective-function-oriented approach to generate alternative clusterings that uses conditional entropy for quantifying both clustering quality and distinctiveness, resulting in an analytically consistent combined criterion. It uses the following objective function:

$$C = argmin\left\{\sum_{k=1}^{K} p(c_k) H_2(X|C = c_k)\right\} \quad (3)$$

where $K$ is the number of clusters and $H_2$ is the Havrda-Charvat quadratic conditional entropy that can be estimated by

$$H_2(X) = 1 - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(x_i - x_j, 2\sigma^2) \quad (4)$$

if one uses the Parzen window method with a Gaussian kernel $G(.)$ to estimate the probability density function. This algorithm has a similar approach as the MEC algorithm by trying to evaluate at each step the change in entropy when assigning a point to a different cluster. Three parameters are needed: the number of clusters for the initial partition, the Parzen window size and the number of steps of the algorithm.

### B. Other Clustering Algorithms

In the following we briefly describe spectral clustering, mean-shift and k-means algorithms used on the comparison against entropic clustering algorithms.

The rationale of spectral clustering is to use special properties of the eigenvectors of a Laplacian matrix as the basis to perform clustering. The Laplacian matrix is based on an affinity matrix, $A$, built with a similarity measure. The most common similarity measure used in spectral clustering

is $A_{ij} = \exp\left(-d_{ij}^2/2\sigma^2\right)$, where $d_{ij}$ is the Euclidian distance between vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ and $\sigma$ is a scaling parameter. With matrix $A$, the Laplacian matrix $L$ is computed as $L = D - A$, where $D$ is the diagonal matrix whose elements are the sums of all row elements of $A$.

There are several spectral clustering algorithms. We use here three different ones:

- The Shi and Malik normalized cut algorithm [8] that perform cutting using the second smallest eigenvector of the "normalized" Laplacian matrix.
- The Ng and Jordan algorithm [9] that uses the highest eigenvectors as input to another clustering algorithm.
- The Perona and Freeman algorithm [10] that is related with Shi and Malik normalized cuts algorithm and uses the concept of affinity factorization to build the clusters.

We also used Mean shift [11] in the comparison. Mean shift is a density based clustering algorithm mostly used on image analysis. It determines density associated with each point by applying a kernel in a neighborhood of that point in order to find local maximums of the density function that are used to build the clusters.

The final algorithm is the well known k-means. K-means is an iterative algorithm that starts by defining a given number of centroids and at each step assigns each point to a given cluster depending on the distance to each centroid. The centroids are then updated until a minimum of a distance based objective function is achieved.

## IV. Datasets and Metrics

### A. Datasets

Since the purpose of this work is to evaluate entropic clustering algorithm on biomedical data we chose 23 publicly available datasets of this kind summarized in Table I. These datasets are usually 2-class problems. The reason is related with the fact that the most common medical and biomedical problems deal with the existence or not of a certain disease. Datasets with more classes are more uncommon and are usually related with several stages of the disease or a certain clinical or health condition.

Datasets Breast Cancer W., Column 2C, Column 3C, Lung Cancer-uci and Spectf can be obtained in the UCI repository, Asthma, Babies, Growth, Surgery, Weights and Xray were obtained from [12], Acath, Diabetes, Dmd, Dominican, Gbsg, Pbc, Prostate and Stress echo can be found in [13] and Depress, Lung cancer-ucla and Lung can be downloaded from [14].

Table I
Datasets characteristics.

| Name | Cases | Features | Classes |
|---|---|---|---|
| Acath | 2258 | 5 | 2 |
| Asthma | 2464 | 3 | 2 |
| Babies | 256 | 3 | 2 |
| Breast cancer W. | 699 | 10 | 2 |
| Column 2C | 310 | 6 | 2 |
| Column 3C | 310 | 6 | 3 |
| Column full | 620 | 6 | 4 |
| Depress | 294 | 10 | 2 |
| Diabetes | 366 | 13 | 3 |
| Dmd | 185 | 5 | 3 |
| Dominican | 318 | 3 | 2 |
| Gbsg | 686 | 7 | 2 |
| Growth | 277 | 6 | 2 |
| Lung cancer-uci | 27 | 102 | 3 |
| Lung cancer-ucla | 327 | 4 | 2 |
| Lung function | 150 | 36 | 4 |
| Pbc | 276 | 13 | 2 |
| Prostate | 483 | 19 | 2 |
| Spectf | 248 | 88 | 2 |
| Stress echo | 558 | 15 | 2 |
| Surgery | 126 | 6 | 2 |
| Weights | 550 | 10 | 4 |
| Xray | 150 | 10 | 2 |

### B. Adjusted Rand Index

There are several performance indexes for cluster evaluation. In this work we use the Adjusted Rand Index (ARI) [15] that measures the correspondence between two partitions of the same data and its based on how pairs of objects are classified in a contingency table. This is a very useful measure when we want to evaluate clustering algorithms and when the labels are known because we can compare the clustering results with the true classes.

Consider a set of $n$ objects $S = \{O_1, O_2, ..., O_n\}$ and suppose that $U = \{u_1, u_2, ..., u_R\}$ and $V = \{v_1, v_2, ..., v_C\}$ represent two different partitions of the objects in $S$ such that $\cup_{i=1}^{R} u_i = S = \cup_{j=1}^{C} v_j$ and $u_i \cap u_{i'} = \emptyset = v_j \cap v_{j'}$ for $1 \leq i \neq i' \leq R$ and $1 \leq j \neq j' \leq C$. Given two partitions, $U$ and $V$, with $R$ and $C$ subsets, respectively, the contingency Table II can be formed to indicate group overlap between $U$ and $V$.

Table II
Contingency Table for Comparing Partitions $U$ and $V$.

| Partition | | $V$ | | | | |
|---|---|---|---|---|---|---|
| | Group | $v_1$ | $v_2$ | $\cdots$ | $v_C$ | Total |
| $U$ | $u_1$ | $t_{11}$ | $t_{12}$ | $\cdots$ | $t_{1C}$ | $t_{1.}$ |
| | $u_2$ | $t_{21}$ | $t_{22}$ | $\cdots$ | $t_{2C}$ | $t_{2.}$ |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| | $u_R$ | $t_{R1}$ | $t_{R2}$ | $\cdots$ | $t_{RC}$ | $t_{R.}$ |
| Total | | $t_{.1}$ | $t_{.2}$ | $\cdots$ | $t_{.C}$ | $t_{..} = n$ |

In Table II, generic entry, $t_{rc}$, represents the number of objects that were classified in the *r*th subset of partition $R$ and in the *c*th subset of partition $C$. Considering the total number of possible combinations of pairs $\binom{n}{2}$ from a given set one can represent them in four different types of pairs: $a$ - objects in a pair are placed in the same group in $U$ and

in the same group in $V$;
$b$ - objects in a pair are placed in the same group in $U$ and in different groups in $V$;
$c$ - objects in a pair are placed in the same group in $V$ and in different groups in $U$ and;
$d$ - objects in a pair are placed in different groups in $U$ and in different groups in $V$.

This leads to an alternative representation of Table II as a $2 \times 2$ contingency table (Table III) based on $a$, $b$, $c$, and $d$.

The values of the four cells in Table III can be calculated using the values of Table II by:

$$a = \sum_{r=1}^{R}\sum_{c=1}^{C}\binom{t_{rc}}{2} = \left(\sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2 - n\right)/2 \tag{5}$$

$$b = \sum_{r=1}^{R}\binom{t_{r.}}{2} - a = \left(\sum_{r=1}^{R} t_{r.}^2 - \sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2\right)/2 \tag{6}$$

$$c = \sum_{c=1}^{C}\binom{t_{.c}}{2} - a = \left(\sum_{c=1}^{C} t_{.c}^2 - \sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2\right)/2 \tag{7}$$

$$\begin{aligned} d &= \binom{n}{2} - a - b - c \\ &= \binom{n}{2} - \sum_{r=1}^{R}\binom{t_{r.}}{2} - \sum_{c=1}^{C}\binom{t_{.c}}{2} + a \\ &= \left(\sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2 + n^2 - \sum_{r=1}^{R} t_{r.}^2 - \sum_{c=1}^{C} t_{.c}^2\right)/2 \end{aligned} \tag{8}$$

where $t_{rc}$ represents each element of the $R \times C$ matrix of Table II.

Table III
SIMPLIFIED $2 \times 2$ CONTINGENCY TABLE FOR COMPARING PARTITIONS $U$ AND $V$.

| Partition $U$ | $V$ | |
|---|---|---|
| | Pair in same group | Pair in different groups |
| Pair in same group | $a$ | $b$ |
| Pair in different groups | $c$ | $d$ |

Several different performance indexes such as the Jaccard [16] and Rand [17] indexes are computed using these four values. To overcome some limitations of these indexes such as the problem that the expected value does not take a constant value some improved measures were created. Examples are the Fowlkes-Mallows [18] Index or the Adjusted Rand Index (ARI) proposed by Hubert and Arabie [15] as an improvement of Rand Index. ARI is in fact recommended as the index of choice for measuring agreement between two partitions in clustering analysis with different numbers of clusters [19]. It can be computed as:

$$ARI_{(U,V)} = \frac{\binom{n}{2}(a+d) - [(a+b)(a+c)+(c+d)(b+d)]}{\binom{n}{2}^2 - [(a+b)(a+c)+(c+d)(b+d)]} \tag{9}$$

or

$$\frac{\binom{n}{2}\sum_{r=1}^{R}\sum_{c=1}^{C}\binom{t_{rc}}{2} - \left[\sum_{r=1}^{R}\binom{t_{r.}}{2}\sum_{c=1}^{C}\binom{t_{.c}}{2}\right]}{\frac{1}{2}\binom{n}{2}\left[\sum_{r=1}^{R}\binom{t_{r.}}{2} + \sum_{c=1}^{C}\binom{t_{.c}}{2}\right] - \left[\sum_{r=1}^{R}\binom{t_{r.}}{2}\sum_{c=1}^{C}\binom{t_{.c}}{2}\right]} \tag{10}$$

with expected value zero and maximum value 1.

## V. RESULTS AND DISCUSSION

In Table IV we present the best results (ARI values) of the performed experiments. This means that we have tuned the parameters for each algorithm in order to obtain them. The best result for each dataset is underlined. In the last column we present the number of best results for each algorithm. There are some negative values in the referred table but this has to do with the estimation of the expected value in ARI index what can lead to small deviations and produce small negative values when the result is very close to zero.

Results show that there is not, as expected, a unique algorithm that produces the best results for all datasets.

The first important thing that one can see on the results is that the ARI values are in general extremely small indicating that the clustering result is very poor. One can infer from this fact that these datasets are highly complex and the existence of groups in the data is very dubious. Usually, the classes in this kind of data are very overlapped and consequently the groups are not well defined making the clustering process much more difficult.

When comparing the results of the clustering algorithms on the 23 datasets, one can see that entropic based clustering algorithms present better results on 15 of them and the other algorithms present better results on 13. In general terms, one may conclude that there is no clear superiority of a specific group of algorithms but entropic clustering algorithms present a slightly better performance. However, when performing an individual comparison, Minimum Entropic Clustering (MEC) presents a clear advantage over the others with a considerable higher number of best results (8). The second bests are LEGClust and k-means, each one with 4 best results (half the number of MEC). Results of spectral clustering are very poor when compared to entropic clustering and even k-means. Mean-Shift algorithm presents the worst results on the performed experiments.

## VI. CONCLUSION

The purpose of this work was to evaluate entropic based clustering algorithms on biomedical data and to see if there was a significant difference on the performance of these

Table IV
RESULTS (ARI INDEX) OF THE DIFFERENT ALGORITHMS FOR THE 23 BIOMEDICAL DATASETS.

| | | | | | Spectral CLustering | | | |
|---|---|---|---|---|---|---|---|---|
| | LEGC | MEC | MCE | k-means | SM | NJ | PF | MS |
| Acath | 0.000 | 0.020 | 0.115 | 0.008 | 0.006 | -0.016 | -0.016 | 0.000 |
| Asthma | 0.016 | 0.021 | 0.042 | 0.020 | -0.004 | 0.042 | 0.042 | 0.021 |
| Babies | 0.080 | -0.004 | 0.018 | -0.004 | 0.001 | -0.039 | 0.000 | 0.000 |
| Breast cancer W. | 0.721 | 0.662 | 0.450 | 0.691 | 0.004 | -0.001 | 0.003 | 0.172 |
| Column 2C | 0.108 | 0.187 | 0.166 | 0.106 | 0.000 | 0.034 | 0.000 | 0.001 |
| Column 3C | 0.270 | 0.251 | 0.217 | 0.275 | 0.001 | 0.149 | 0.000 | 0.008 |
| Column full | 0.152 | 0.116 | 0.141 | 0.126 | 0.007 | 0.058 | 0.000 | 0.047 |
| Depress | 0.035 | -0.007 | 0.041 | -0.004 | 0.000 | -0.002 | 0.061 | 0.036 |
| Diabetes | 0.120 | 0.063 | 0.054 | 0.065 | 0.000 | 0.004 | 0.001 | 0.000 |
| Dmd | 0.085 | 0.138 | 0.127 | 0.199 | 0.010 | 0.009 | -0.002 | 0.019 |
| Dominican | 0.014 | -0.009 | -0.008 | -0.009 | 0.017 | -0.002 | 0.003 | -0.002 |
| Gbsg | 0.005 | 0.045 | 0.043 | 0.043 | 0.000 | 0.026 | 0.000 | 0.000 |
| Growth | 0.070 | 0.087 | 0.007 | 0.066 | 0.000 | 0.003 | 0.000 | 0.028 |
| Lung cancer-uci | 0.219 | 0.224 | 0.026 | 0.224 | 0.296 | 0.296 | 0.074 | -0.004 |
| Lung cancer-ucla | -0.003 | 0.103 | 0.0072 | 0.007 | 0.004 | 0.007 | 0.007 | 0.143 |
| Lung function | 0.028 | 0.035 | 0.021 | 0.026 | 0.005 | 0.095 | 0.000 | -0.008 |
| Pbc | 0.174 | 0.273 | 0.253 | 0.273 | 0.025 | 0.025 | 0.014 | 0.026 |
| Prostate | 0.006 | 0.527 | 0.006 | 0.000 | 0.014 | 0.001 | 0.003 | -0.003 |
| Spectf | 0.007 | -0.103 | -0.086 | -0.103 | 0.023 | 0.003 | -0.003 | -0.081 |
| Stress echo | 0.003 | 0.004 | -0.006 | -0.022 | 0.001 | -0.002 | -0.001 | -0.019 |
| Surgery | 0.001 | 0.025 | 0.019 | 0.025 | 0.000 | 0.004 | -0.004 | -0.011 |
| Weights | 0.006 | 0.013 | 0.004 | 0.001 | 0.001 | 0.000 | 0.000 | -0.005 |
| Xray | 0.000 | -0.008 | 0.341 | -0.007 | -0.010 | 0.320 | 0.320 | -0.030 |
| Best results | 4 | 8 | 3 | 4 | 3 | 3 | 2 | 1 |

LEGC - LEGClust; MEC - Minimum Entropic Clustering; MCE - minCEntropy;
SM - Shi & Malik; NJ - Ng & Jordan; PF - Perona & Freeman; MS - Mean-Shift

algorithms when compared with other ones that are usually considered in the literature.

Results show that the difference is not very significative when evaluating the results on general terms but when considering each algorithm individually there is a superior performance of Minimum Entropic Clustering algorithm. Since the number of parameters to tune in MEC is 3 and in k-means we only need the number of clusters one should weight the pros and the cons when deciding which one to use in the future.

## REFERENCES


[1] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[2] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, no. 2-3, pp. 107–145, 2001.

[3] D. Barbará, Y. Li, and J. Couto, "Coolcat: an entropy-based algorithm for categorical clustering," in *Proceedings of the Eleventh International Conference on Information and Knowledge Management*. ACM, 2002, pp. 582–589.

[4] N. Karayiannis, "MECA: Minimum entropy clustering algorithm," in *IEEE World Congress on Computational Intelligence*, vol. 1, 1994, pp. 630 – 635.

[5] H. Li, K. Zhang, and T. Jiang, "Minimum entropy clustering and applications to gene expression analysis," in *Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference*. IEEE Computer Society, 2004, pp. 142–151.

[6] J. M. Santos, J. Marques de Sá, and L. A. Alexandre, "LEGClust - a clustering algorithm based on layered entropic subgraphs," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006, submitted.

[7] N. Vinh and J. Epps, "MinCEntropy: A novel information theoretic approach for the generation of alternative clusterings," in *IEEE International Conference on Data Mining*, 2010, p. employ the conditional entropy measure for quantifying both clustering quality and distinctiveness.

[8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[9] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, 2001.

[10] P. Perona and W. Freeman, "A factorization approach to grouping," in *Lecture Notes in Computer Science*, Springer, Ed., no. 1406, 1998, pp. 655–670.

[11] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[12] J. Peat and B. Barton, "Medical statistics: A guide to data analysis and critical appraisal," http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0727918125

[13] F. Harrell, http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets.

[14] A. Afifi, "Biostatistics 406 datasets," http://www.biostat.ucla.edu/course/406/datasets.htm.

[15] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[16] P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," in *Bulletin del la Société Vaudoise des Sciences Naturelles*, 1901, no. 37, pp. 547–579.

[17] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, pp. 846–850, 1971.

[18] E. Fowlkes and C. Mallows, "A method for comparing two hierarchical clusterings," *Journal of the American Statistical Association*, vol. 78, pp. 553–569, 1983.

[19] G. Milligan and M. Cooper, "A study of the comparability of external criteria for hierarchical cluster analysis," *Multivariate Behavioral Research*, vol. 21, pp. 441–458, 1986.