OVERVIEW OF AGGLOMERATIVE HIERARCHICAL CLUSTERING METHODS

Eric U. Oti[1*] and Michael O. Olusola[2]

[1]Department of Statistics, Federal Polytechnic, Ekowe, Bayelsa State.

[2]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State.

[*]Corresponding Author: eluchcollections@gmail.com; Tel.: +2348037979262

Cite this article:

Oti, Eric U., Olusola, Michael O. (2024), Overview of Agglomerative Hierarchical Clustering Methods. British Journal of Computer, Networking and Information Technology 7(2), 14-23. DOI: 10.52589/BJCNIT-CV9POOGW

Manuscript History

Received: 14 Mar 2024

Accepted: 6 May 2024

Published: 7 Jun 2024

ABSTRACT: *Agglomerative hierarchical clustering methods are the most popular type of hierarchical clustering used to group objects in clusters based on their similarity. The methods uses a bottom-up approach and it starts clustering by treating the individual data points as a single cluster, then it is merged continuously based on similarity until it forms one big cluster containing all objects. In this paper, we reviewed eight agglomerative hierarchical clustering methods namely: single linkage method, complete linkage method, average linkage method, weighted group average method, centroid method, median method, Ward's method and the flexible beta method; we also discussed measures of similarity and dissimilarity using quantitative data as our reference point.*

KEYWORDS: Cluster, Dendrogram, Dissimilarity, Objects, Similarity.

INTRODUCTION

Clustering in statistics is purely seen as a multivariate technique but can also be applied to univariate and bivariate data. Furthermore, it is a multivariate technique where a set of data, usually multidimensional is classified into clusters such that members of one cluster are similar to one another with respect to some predetermined criterion (Gan et al. 2007; Everitt et al. 2011). The clusters of objects should exhibit high internal homogeneity and high external heterogeneity (MacQueen, 1967; Anderberg, 1973). Clustering is done on the basis of similarities or dissimilarities (Kaufman and Rousseeuw, 1990; Johnson and Wichern, 2002; Mirkin, 2013).

Clustering methods can be widely classified into two main groups which are based on the structure of the output namely: hierarchical and non-hierarchical clustering methods. Hierarchical clustering is a method of cluster analysis that seeks to build a hierarchy of clusters, and it is classified into two main parts namely: agglomerative and divisive methods. In hierarchical clustering, clusters are merged (agglomerative methods) and split (divisive methods) step by step based on applied similarity measure. The outcome of a hierarchical clustering method suffice that agglomerative and divisive methods can be displayed graphically using a tree diagram called dendrogram, while non-hierarchical clustering methods partition dataset into clusters where every pair of object clusters is either distinct or has some members in common. Agglomerative hierarchical clustering has been the dominant approach to constructing embedded classification schemes. It is our aim to direct the reader’s attention to practical methods that are both effective and efficient.

The purpose of this paper is to present a general survey of agglomerative hierarchical clustering methods and its measure of similarity and dissimilarity.

The rest of this paper is organized as follows: Section 2 discussed the measures of similarities and dissimilarities (distance) of agglomerative clustering using quantitative data. Section 3 discussed briefly on related literature. Section 4 discussed the methods of agglomerative hierarchical clustering methods and Section 5 is the conclusion of the paper.

MEASURES OF SIMILARITY AND DISSIMILARITY

Clusters are considered as groups containing data objects that are similar to each other than data objects in different clusters. Thus, in attempting to identify clusters of observations which may be present in data is knowledge on how “close” individuals or objects are to each other, or how far apart they are from each other (Jain and Dubes, 1988; Xu and Wunsch, 2008). Many clustering investigations have as their starting point a one-mode matrix, the elements which reflect in some sense, a quantitative measure of closeness, commonly referred to as dissimilarity (distance) or similarity (Oti and Olusola, 2024), with a general term being known as proximity. Two individuals or objects are “close” when their dissimilarity is small or their similarity is large (Everitt, et al., 2011; Romesburg, 1984).

The term proximity is the generalization for both dissimilarity and similarity. A dissimilarity or distance function on a data set x is defined to satisfy the following condition of a metric space (Anderberg, 1973; Zhang and Srihari, 2003):

Article DOI: 10.52589/BJCNIT-CV9POOGW
DOI URL: https://doi.org/10.52589/BJCNIT-CV9POOGW

- $d(x,y) \geq 0 \; for \; all \; x \; and \; y$ (Non-negativity)
- $d(x,y) = d(y,x)$ (Symmetry)
- $d(x,y) \leq d(x,z) + d(z,y) \; for \; all \; x,y,z$ (Triangle inequality)
- $d(x,y) = 0 \; iff \; x = y$ (Reflexivity)

A metric space (X,d) is a set X with a metric d defined on X, but if the triangle inequality is not satisfied, the function is called a semi-metric. While a metric is a function that defines a concept of distance between any two points of the set; and also, if a metric is an ultra-metric (Johnson, 1967) implies that it satisfies a stronger condition that states that: $d(x,y) \leq max\,\{d(x,y)\} \; for \; all \; x,y,z$ where x,y,z are arbitrary data points.

In hierarchical clustering, a range of measures calculate either the similarities or dissimilarities associated with two pairs of observations $x \; and \; y$. The word similarity measure is a way of measuring how pair of observations is closer to each other. While dissimilarity measure tells us how different pair of observations is from each other.

To group data, we need a way to measure the elements and their distance relative to each other in order to decide which elements belong to a group. This can be a similarity, although on many occasions a dissimilarity measurement is used (Murtagh and Contreras, 2011). In application, the choice of distance is important and the best choice is often achieved through the combination of experience, knowledge, skill and sometime luck. The most common used distance for quantitative data is the Euclidean distance. Euclidean and the Manhattan distances are the special cases of the Minkowski distance defined as:

$$\mathrm{d(x,y)} = \left(\sum_{\mathrm{J}=1}^{\mathrm{d}} \left\lfloor \mathrm{x_j} - \mathrm{y_j} \right\rfloor^{\mathrm{r}}\right)^{\frac{1}{\mathrm{r}}} \quad (1)$$

Where r is called the power of the Minkowski distance; when r = 2 and 1, we get the Euclidean and Manhattan distance which are found in Equation (2) and Equation (3) below:

$$\mathrm{d(x,y)} = \left(\sum_{\mathrm{j}=1}^{\mathrm{d}} (\mathrm{x_j} - \mathrm{y_j})^2\right)^{\frac{1}{2}} = \left((\mathrm{x}-\mathrm{y})^{\mathrm{T}}(\mathrm{x}-\mathrm{y})\right)^{\frac{1}{2}} \quad (2)$$

$$\mathrm{d(x,y)} = \sum_{\mathrm{j}=1}^{\mathrm{d}} \left|\mathrm{x_j} - \mathrm{y_j}\right| \quad (3)$$

Other distances used in quantitative data are:

Chi-square distance is one of the distance measures that can be used as a measure of dissimilarity between two histograms and has been widely used in various applications such as image retrieval, feature extractions, image texture and object classification. Chi-square distance measures similarity between two feature matrices. The Chi-square distance of 2 arrays x and y with n dimension is mathematically calculated the formula below:

$$d(x, y) = \sum_{j=1}^{n} \frac{\left(x_j - y_j\right)^2}{\left(x_j + y_j\right)} \qquad (4)$$

Pearson's dissimilarity is a transformation of the linear (Pearson's r) correlation between two vectors. The linear correlation (Pearson's r) is computed as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (5)$$

When used as a dissimilarity measure, it is rescaled to the interval ⟦0,1⟧ with 0 indicating perfect similarity (perfect positive correlation) and 1 indicating perfect dissimilarity (perfect negative correlation) and can be computed as:

$$d_r(x, y) = \frac{(1 - r)}{2} \qquad (6)$$

Where r is the linear correlation.

Rank dissimilarity is a transformation of Spearman's non-parametric r_s correlation between two vectors. Correlation (Spearman's Rank coefficient) is computed as:

$$r_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}} \qquad (7)$$

Where R_i is the rank of x_i in the vector x, S_i is the rank of y_i in the vector y. When used as a dissimilarity measure, it is rescaled to the interval ⟦0,1⟧ with 0 indicating perfect similarity (perfect positive correlation) and 1 indicating perfect dissimilarity (perfect negative correlation).

$$d_{r_s}(x, y) = \frac{1 - r_s(x,y)}{2} \qquad (8)$$

Where r_s is the Spearman's rank order coefficient.

Kendall's dissimilarity is dissimilarity metric based on the correlation between the vectors and is computed as:

$$d_\tau(x, y) = \frac{(1 - \tau)}{2} \qquad (9)$$

Where τ is Kendall's Tau correlation. It is rescaled to the interval ⟦0,1⟧ with 0 indicating perfect similarity (perfect positive correlation) and 1 indicating perfect dissimilarity (perfect negative correlation).

Measures of similarity as regards quantitative data looking at the most common way of measuring a linear correlation is attributed to the Pearson's coefficient of correlation (r) which represents the relationship between two variables denoted as x and y and the measure of similarity ranges from +1 to -1. The Spearman's coefficient of rank correlation is another measure of similarity, although it is a nonparametric measure of rank correlation which is often denoted by the Greek letter ρ(rho) which is primarily used for data analysis. The Kendall's coefficient of rank correlation evaluates the degree of similarity between two sets of ranks given to the same set of objects. Kendall's rank correlation is also a nonparametric and it is an alternative to Pearson's correlation (parametric) when the data you are working with has failed one or more assumptions of the test. This is also the best alternative to Spearman correlation when your sample size is small and has many tied ranks.

RELATED WORK

Podani (1989) proposed a method of combinatorial SAHN (sequential, agglomerative, hierarchical and non-overlapping) classificatory strategies which are subdivided into two classes: The d-SAHN method which seeks for minimal between-cluster distances, while the h-SAHN strategies for maximal within-cluster homogeneity. The method described a weighted and unweighted variant of the minimization of the increase of average distance within clusters and a homogeneity-optimizing flexible method.

Ah-Pine (2018) proposed a method called sparsified normalized kernel matrix based agglomerative hierarchical clustering method which framework is generic, efficient and effective. The approach embeds a sub-family of Lance-Williams (LW) clustering which relies on inner-products instead of squared Euclidean distance.

Mohammad et al. (2022) proposed an agglomerative hierarchical clustering method that depends on ensemble-based approach. The proposed algorithm consists of three main steps: In the first step, a group of single agglomerative hierarchical clustering methods are combined to detect relationships between samples and as well as the formation of initial clusters. The similarity of the samples is calculated using an innovative similarity criterion based on the clusters created. In the second step, all the initial clusters created by different methods are re-clustered to form hyper-clusters. After cluster clustering each sample is assigned to a hyper-cluster with maximum similarity to create the final clusters in the third step.

AGGLOMERATIVE HIERARCHICAL CLUSTERING METHODS

The agglomerative clustering is the most common type of hierarchical clustering method used in grouping objects in clusters based on their similarity (Everitt et al., 2011). The algorithm starts by treating each object as a singleton cluster. Then, it repeats merging the closest pair of clusters according to some similarity criteria until all of the data are in one cluster (Gan et al., 2007). According to different distance measures between groups, agglomerative hierarchical methods can be subdivided into Single-linkage method, Complete-linkage method, Average-linkage method, Weighted group average method, Centroid method, Median method, Ward's method and Flexible-Beta method.

The Single-linkage method is one of the simplest agglomerative hierarchical clustering techniques which was first introduced by Florek et al. (1951) and then independently by McQuitty (1957) and Sneath (1957). The single-linkage method is also known by other names such as the nearest neighbor method, the minimum method, or the connectedness method (Rohif, 1982). This method utilizes a minimum distance rule that start by first, the two objects having the shortest (smallest) or largest similarity distance is merged; they constitute the first cluster. At the next stage, one of these two things can happen: Either a third object will join the already formed cluster of two, or the two closest un-clustered objects are joined to form the second cluster. The decision rests on whether the distance from one of the un-clustered objects to the first cluster is shorter than the distances between the two closest un-clustered objects. The process continues until all objects belong to a single cluster. To find the minimum distance in $D = \{d_{ik}\}$, merge the corresponding objects: say x and y to get the cluster (xy) and any other cluster z are computed as $d_{(XY)Z} = min\{d_{XZ}, d_{YZ}\}$, where d_{XZ}, and d_{YZ} are the distance between the nearest neighbors of cluster x and y respectively.

The complete linkage method (McQuitty, 1960; Sokal and Sneath, 1963) uses the farthest neighbor distance to measure the dissimilarity between two groups. This method ensures that all items in a cluster are within the same maximum distance (or minimum similarity) to each other. To find the maximum distance in $D = \{d_{ik}\}$, we merge the corresponding objects like x and y to get cluster (xy). For the next step, the distance between clusters (xy) and any other cluster z are computed as $d_{(XY)Z} = max\{d_{XZ}, d_{YZ}\}$, where d_{XZ} and d_{YZ} are distances between the most distance members of the cluster x and z and cluster y and z respectively. This method begins by searching the distance matrix $D = \{d_{ik}\}$ to find the nearest (most similar) object x and y. Objects are merged to form cluster (xy). Update the entries in the distance matrix by deleting the rows and columns corresponding to cluster x and y by adding a row and column giving the distance between cluster (xy) and the remaining clusters. The process is repeated until all objects are in a single cluster after the algorithm terminates. Record the identity of clusters that are merged and the levels (distances or similarities) at which the mergers take place.

The average linkage method proposed by Sokal and Michener (1958) is sometimes referred to as un-weighted pair group method using arithmetic averages (Jain and Dubes, 1988). This method treats the distance between two clusters as the average distance between all pairs of items where one member of the pair belongs to each cluster. The distance can be computed as

$$d_{(XY)Z} = \frac{\sum_i \sum_k d_{ik}}{N_{(xy)} N_z} \qquad (10)$$

Where d_{ik} is the distance between object i in the cluster (xy) and object k in the cluster z; $N_{(xy)}$ and N_z are the number of items in cluster (xy) and z respectively.

The weighted group average method proposed by McQuitty (1966) is also referred to as weighted pair group method using arithmetic average (WPGMA) is similar to average linkage method but weights inter-cluster distances according to the inverse of the number of objects in each class, as in the case of median compared to centroid linkage method (Everitt et al. 2011). The method constructs a rooted dendrogram that reflects the structure present in a pairwise distance or similarity matrix. At each step, the nearest two clusters i and j are combined into a

Article DOI: 10.52589/BJCNIT-CV9POOGW
DOI URL: https://doi.org/10.52589/BJCNIT-CV9POOGW

higher level cluster $i \cup j$, then its distance to another cluster k is simply the arithmetic mean of the average distances between members of k and i and k and j is computed by:

$$d_{(i \cup j),k} = \frac{d_{i,k} + d_{j,k}}{2} \tag{11}$$

The centroid method proposed by Sokal and Michener (1958) is a method where the distance between two clusters i and k is defined as the Euclidean distance between the mean vectors (often called centroids) of the two clusters are stated as:

$$d(i,k) = d(\bar{x}_i \bar{x}_k) \tag{12}$$

Where $\bar{x}_i$ and $\bar{x}_k$ are the mean vectors for the observation in i and the observation in k respectively, while $d(\bar{x}_i, \bar{x}_k)$ is the respective distance mean in the Euclidean space. The two clusters with the smallest distance between centroid are merged at each step; after two clusters i and k are joined; the centroid of the new cluster (ik) is calculated using the weighted average

$$\bar{x}_{ik} = \frac{n_i \bar{x}_i + n_k \bar{x}_k}{n_i + n_k} \tag{13}$$

The centroid method is sometimes referred as un-weighted pair group method using centroid.

The median method also known as the "weighted pair group method using centroid" (Jain and Dubes, 1988) was proposed by Gower (1967) in order to alleviate some disadvantages of the centroid method. In the centroid method, if the sizes of the two groups to be merged are quite different, then the centroid of the new group will be very close to that of the large group and may remain within that group (Everitt, 1993). While in the centroid method, the centroid of the new group is independent of the size of the groups that form the new group. A disadvantage of this method is that it is not suitable for measures such as correlation coefficients, since interpretation in a geometrical sense is no longer possible (Lance and Williams, 1967). To avoid weighting the mean vector according to cluster size, we can use the median (midpoint) of the line joining i and k as the point for computing next distance to other clusters $M_{ik} = \frac{1}{2}(\bar{x}_i + \bar{x}_k)$. The two clusters with the smallest distance between median are merged at each step.

The wards method was proposed by Ward and Hook (1963), it is a procedure seeking method that forms the partition $p_k, p_{k-1}, p_{k-2}, \dots, p_1$ in a manner that minimizes the loss of information associated with the merging. Usually, the loss of information is quantified in terms of error sum of squares (ESS) criterion, so Ward's method is often referred to as the "minimum variance" method. Given a group c, the associated with c is given by: $ESS(c) = \sum_{x \in c} (x - \mu(c))(x - \mu(c))^T = \sum_{x \in c} xx^T - \frac{1}{|c|}(\sum_{x \in c} x)(\sum_{x \in c} x)^T \quad = \sum_{x \in c} xx^T - |c|\mu(c)\mu(c)^T$ Where $\mu|c|$ is the mean of c, that is, $\mu|c| = \frac{1}{|c|}\sum_{x \in c} x$. Suppose there are k groups $c_1, c_2, \dots, c_k$ in one level of the clustering, then the information loss is represented by $ESS = \sum_{i=1}^{k} ESS(c_i)$ which is the total within-group of the ESS. At each step of Ward's method, the union of every possible pair of groups is considered and two groups whose fusion results in the minimum

Article DOI: 10.52589/BJCNIT-CV9POOGW
DOI URL: https://doi.org/10.52589/BJCNIT-CV9POOGW

increase in loss of information are merged. If the squared Euclidean distance is used to complete the dissimilarity matrix, then the dissimilarity matrix can be updated by the Lance-William formula (Wishart, 1969) during the process of clustering as follows:

$$D\left(c_k, c_i \cup c_j\right) = \frac{|c_k|+|c_i|}{\Sigma_{ijk}} D(c_k, c_i) + \frac{|c_k|+|c_j|}{\Sigma_{ijk}} D\left(c_k, c_j\right) - \frac{|c_k|}{\Sigma_{ijk}} D\left(c_i, c_j\right) \qquad (14)$$

Initially, each single point forms a cluster and the total ESS is $ESS_k = 0$.

The Flexible-Beta technique was proposed by Lance and Williams (1967). In this technique, suppose cluster A and B are merged to form cluster AB. A general formula for the distance between AB and any other cluster C is given as:

$$D(C, AB) = \propto AD(C, A) + \propto BD(C, B) + BD(A, B) + Y|D(C, A) - (C, B)| \qquad (15)$$

The distances $D(C, A). D(C, B), D(A, B)$ are from the distance matrix before joining A and B. The distance from AB to the other clusters as given by Equation (15) would be used (along with distance between other pairs of clusters) to form the next distance matrix for clustering the pair of clusters with smallest distance, which pair would then be joined at the next step. To simplify Equation (15), Lance and Williams (1967) suggested the following constraints on the parameter values:

$$\propto A + \propto \beta + \beta = 1$$

$$\propto A = \propto \beta$$

$$\gamma = 0$$

$$\beta > 0$$

With $\propto A = \propto \beta$ and $y = 0$, we have $2 \propto a = 1 - \beta$ or $\propto A = \propto \beta = (1 - \beta)/2$ and we need only choose a value of β. The choice of β determines the characteristics of the flexible-Beta clustering procedure. Lance and Williams (1967) suggested the use of a small negative value of β, such as $\beta = -0.25$ if there are outliers in the data, the use of a smaller value of β, such as $\beta = -0.5$ may be more suitable to isolate these outliers into simple clusters.

CONCLUSION

In this paper, we have reviewed the eight agglomerative hierarchical clustering methods and the measures of similarity and dissimilarity using quantitative data as our reference point. The dendrograms of agglomerative hierarchical clustering will always show all the steps in the hierarchical procedure which will include the distances or similarities at which clusters are merged.

ACKNOWLEDGEMENT

This research has been partly funded by the Nigerian Tertiary Education Trust Fund (TETFund), under grant number TETF/DR&D/CE/POLY 1/EKOWE/IBR/2024/VOL. 1.

REFERENCES

Ah-Pine, J. (2018). An efficient and effective generic agglomerative hierarchical clustering approach, Journal of Machine Learning Research, 19(42): 1-43.

Anderberg, M. R. (1973). Cluster Analysis for Applications. Academic Press, New York.

Everitt, B. S. (1993). Cluster Analysis, 3rd Edition. New York, Toronto: Halsted Press.

Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2011). Cluster Analysis, 5th Edition. John Wiley and Sons.

Florek, K., Lukaszewicz, J., Steinhaus, H. and Zubrzycki, S. (1051). Sur la liaison et la division des points d'un ensemble fini. Colloquium Mathematicum. 2:282-285.

Gan, G., Ma, C., and Wu, J. (2007). Data Clustering: Theory, Algorithms, and Applications. ASA-SIAM Series.

Gower, J. C. (1967). Multivariate Analysis and Multidimensional Geometry. Journal of the Royal Statistical Society. Series D (The Statistician) 17(1): 13-28.

Jain, A. and Dubes, R. (1988). Algorithms for Clustering Data. Eaglewood Cliffs, NJ: Prentice-Hall.

Johnson, R. A. and Wichern, D. W. (2002). Applied Multivariate Statistical Analysis. 5th Edition, Eaglewood Cliffs. NJ: Prentice Hall.

Johnson, S. C. (1967). Hierarchical clustering scheme. Psychometrica, 32: 241-254.

Kaufman, L. and Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, Inc.

Lance, G. N. and Williams, W. T. (1967). A general theory of classificatory sorting strategies: Hierarchical systems. The Computer Journal 9(4): 373-380.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp.281-297. Berkeley, CA: University of California Press.

McQuitty, L. (1957). Elementary linkage analysis for isolating orthogonal and oblique types and relevancies, Educational and Psychological Measurement, 17: 207-222.

McQuitty, L. (1966). Similarity analysis by reciprocal pairs for discrete and continuous data. Educational and psychological measurement, 26:825-831.

Mirkin, B. (2013). Clustering: A Data Recovery Approach, Second Edition (Chapman and Hall/CRCComputer Science and Data Analysis).

Mohammad, J., Faramarz, S. and Zahra, B. (2022). An agglomerative hierarchical clustering framework for improving the ensemble clustering process, Cybernetics and Systems 53(3): 1-23, DOI: 10.1080/01969722.2042917

Murtagh, F. and Contreras, P. (2011). Methods of hierarchical clustering. https://doi.org/10.48550/arXiv.1105.0121

Oti, E. U. and Olusola, M. O. (2024). Comparative evaluation of six agglomerative hierarchical clustering methods with a robust example. African Journal of Mathematics and Statistics Studies 7(2): 1-25, DOI: 10.52589/AJMSS-QXPH8R1N

Podani, J. (1989). New combinatorial clustering methods. Vegetatio, 81:61-77.

Rohif, F. (1982). Single link clustering algorithms. In Krishnaiah, P. and Kanal, L., editors. Handbook of Statistics, volumn 2, pages 267-284. Amstadam:North-Holland.

Romesburg, C. (1984). Cluster Analysis for Researchers. London: Wadesworth.

Sneath, P. H. A. (1957). The application of computers to taxonomy. Journal of General Microbiology, 17, 201-226.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin, 38, 1409-1438.

Sokal, R. R. and Sneath, P. H. A. (1963). Principles of Numerical Taxonomy. San Francisco: Freeman.

Ward, Jr., J. and Hook, M. (1963). Application of a hierarchical grouping procedure to a problem of group profiles. Educational and psychological measurement, 23(1): 69-81.

Wishart, D. (1969). 256 Note: An algorithm for hierarchical classifications Biometrics, 25(1): 165-170.

Xu, R. and Wunch, D. C. (2008). Clustering, IEEE Computational Intelligence Society, Series Press. John Wiley and Sons.

Zhang, B. and Srihari, S. (2003). Properties of Binary Vector Dissimilarity Measures Technical Report, CEDAR. Department of Computer Science and Engineering, University of Buffalo, the State University of New York. http://www.cedar.buffalo.edu/papers.html.