An Adaptive Method for Clustering by Fast Search-and-Find of Density Peaks

[Adaptive-DP]

Shanshan Ruan
College of Information Science and Technology
Beijing Normal University
Beijing 100875, China
shanshan_ruan@mail.bnu.edu.cn

Rashid Mehmood*
College of Information Science and Technology
Beijing Normal University
Beijing 100875, China
UMSIT, Kotli, Pakistan
gulkhan007@gmail.com

Ali Daud
King Abdulaziz University, Jeddah, Saudi Arabia
International Islamic University,Isb., Pakistan
adfmohamad@kau.edu.sa

Hussain Dawood
Faculty of Computing and Information Technology
University of Jeddah
Jeddah, Saudi Arabia
hdaoud@uj.edu.sa

Jalal S. Alowibdi
Faculty of Computing and Information Technology
University of Jeddah
Jeddah, Saudi Arabia
jalowibdi@uj.edu.sa

ABSTRACT

Clustering by fast search and find of density peaks (DP) is a method in which density peaks are used to select the number of cluster centers. The DP has two input parameters: 1) the cutoff distance and 2) cluster centers. Also in DP, different methods are used to measure the density of underlying datasets. To overcome the limitations of DP, an Adaptive-DP method is proposed. In Adaptive-DP method, heat-diffusion is used to estimate density, cutoff distance is simplified, and novel method is used to discover exact number of cluster centers, adaptively. To validate the proposed method, we tested it on synthetic and real datasets, and comparison are done with the state of the art clustering methods. The experimental results validate the robustness and effectiveness of proposed method.

Keywords

Clustering; Kernel density estimation; Heat equation; Decision graph

1. INTRODUCTION

Clustering is a technique to organize data points in a way that data points from the same class are more related to one another than the data points in other classes. Clustering is done using different statistical techniques and is widely used in data mining[1, 2, 3], machine learning[4, 5], pattern recognition[6], image analysis[7, 8], cyber security[9], social networks[10, 11], astronomy[12], health care[13], and bioinformatics[14, 15, 16, 17, 18] etc.

Data can be cluster by utilizing different clustering algorithms however; they may differ significantly in what make a cluster and how efficiently clusters are identified. For example, k-means [19] partitions the n points into k classes where each point goes to the class with the minimum mean value, serving as an exemplar of the cluster. However, the k-means has some limitations: the parameter k is hard to assess without external constraints, it is not sensible to noise, and could not detect arbitrary form of clusters.

In density-based methods, clusters are identified as the highest dense regions in the underlying dataset. Data in the sparse regions is mostly considered as noise or border points.

DBSCAN is a basic density-based clustering method [20]. As compared with many other methods, it features the density-reachability cluster model. Like the linkage-based clustering, it finds connecting points within the given radius distance. However, the points were connected based on a certain criteria of connectivity. Also it drops some density points at border regions. Moreover, it could not organize the clusters with overlapping densities[21].

In data mining and statistics, affinity propagation (AP) [22] organizes clusters based on the concept of "message passing" between data points. Different from other clustering algorithms like k-medoids or k-means [23], AP does not required the knowledge of clusters in the data. Similar to k-medoids, AP finds "exemplars" that are representative of clusters.

Clustering by fast search-and-find of density peaks (D-P) was proposed by Alex, et al.[24]. The DP is based on the assumptions that a cluster center is high dense point as compared with its neighbors and located comparatively at higher distance from other cluster centers. For every given data point i, DP estimates it's density ρ_i and distance

*Corresponding author

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054148

δ_i. The effectiveness of DP algorithm highly depends on the evaluation of ρ and cutoff distance d_c. The essential parameter d_c is utilized to estimate the densities, define border points and noise. To identify cluster centers, DP uses the heuristic approach of a decision graph. Users are prompted to identify cluster centers with the paradigm of decision graph. However, organization of clusters created by DP is highly dependent on the d_c, density estimation method, and number of cluster centers selected over the decision graph. The selection of the density estimation method, d_c, and the number of cluster centers are potential barriers to adaptive and effective analysis of data.

To overcome the aforementioned problems, we propose a new method called Adaptive-DP. The Adaptive-DP algorithm uses a heat-diffusion method to estimate density, the selection of d_c is improved, and an adaptive method of selecting cluster centers is introduced. The rest of this paper is organized as follow. Background knowledge is given in Section 2. Section 3 describes the proposed Adaptive-DP in detail. Detailed experimental results and comparisons are given and discussed in Section 4, and finally, the concluding remarks are presented in Section 5.

2. BACKGROUND

Our proposed technique is an extension of DP clustering algorithm. Unlike other clustering strategies, the DP algorithm can define anomalous clusters. Related algorithms such as k-means assume that the clusters are "balls" in given space. The DP algorithm assumes that the center of a cluster has always-higher density as compared to its neighboring points and that the cluster center is comparatively at far distance from other cluster centers. Based on this assumption, the DP computes two quantities for each data point: its local density, and the distance to its nearest high density point. The Algorithm 1 is used to measure the density.

Algorithm 1 Density estimation

Require: d_c, the cutoff distance
$D, n*n$ distance matrix
Ensure: ρ, n length density vector
for $i \leftarrow 1{:}n$ **do**
  $\rho(i) \leftarrow Count(D(i, other_Objects) < d_c)$
end for

According to the algorithm 1, local density ρ_i can be obtained easily, with this quantity and the all-pair distance matrix D as input parameters, the following algorithm 2 is used to get another essential quantity δ_i.

The high dense data points have maximum values of δ. In this way, the data points with higher ρ and higher δ in contrast with other points in the dataset are identified as cluster centers. After computing the two quantities for each data point, the DP plots the calculated values of ρ and δ on a decision graph as shown in Fig.1. Figure 1(a) shows the 28 data points embedded in 2D space in decreasing density order. It is clear that points 1 and 10 are the density maxima; they are thus identified as cluster centers. Figure 1(b) plots the corresponding decision graph using calculated values of ρ and δ. From this representation, it is clear that point 1 and point 10 have higher values of ρ and δ than the other points. The isolated points in Figure 1(a), points 26, 27 and 28, can be found in Figure 1(b) with low ρ and high ρ are treated as noise or outliers. According to the decision graph, the cluster center is the point with high ρ and high δ. After successfully identifying the cluster centers, each remaining data item gets the cluster label based on their δ values in a single round. Algorithm 3 is used to assign points to cluster centers and also to detect noise.

Algorithm 2 Distance from higher local density points (δ_i)

Require: D, $n*n$ Distance matrix; ρ, n size density vector
Ensure: δ,$NNdistance$ vector of n objects from nearest higher density; $NNneighbor$, index vector of nearest neighbor of each element i
for $i \leftarrow 2{:}n$ **do**
  $\delta(sorted\_\rho(i)) \leftarrow \max(D)$
  for $j \leftarrow 1{:}n-1$ **do**
    if $D(sorted\_\rho(i), sorted\_\rho(j)) < \delta(sorted\_\rho(i))$ **then**
      $\delta(sorted\_\rho(i)) \leftarrow D(sorted\_\rho(i), sorted\_\rho(j))$
      $NNneighbor(sorted\_\rho(i)) \leftarrow sorted\_\rho(j)$
    end if
  end for
end for

Algorithm 3 Cluster assignment algorithm

Require: X, Cluster centers; $sorted\_\rho$, density vector of point i, sorted in descending order
Ensure: C, Organized clusters
for $i \leftarrow 1{:}size(X)$ **do**
  $C(i) \leftarrow X(i)$
end for
for $j \leftarrow 1{:}n$ **do**
  if $C(sorted\_\rho(j)) >=' label\_not\_assigned'$ **then**
    $C(sorted\_\rho(j)) \leftarrow C(NNneighbor(sorted\_\rho(j)))$
  end if
end for

3. PROPOSED TECHNIQUE

DP detects density peaks to find the cluster centers. Algorithm 2 or density estimation methods [25, 26] are proposed to estimate the density of underlying datasets [24]. The selection of an appropriate method is purely based on the nature of the underlying dataset. In the proposed method, heat diffusion is used to estimate density in a robust way. Heat diffusion proved to be robust to estimate the density [27, 28, 29]. The densities using the heat diffusion process can be presented by the Eq.1.

$$\hat{f}\,(d;t) = \frac{1}{\mathrm{n}} \sum_{j=1}^{n} \sum_{k=-\infty}^{\infty} e^{-k^2\pi^2 t/2} \cos(k\pi d)\,\cos(k\pi d_j)\ , \tag{1}$$

Equation 1 can be express as

$$\hat{f}\,(d;t) \approx \sum_{k=0}^{n-1} a_k e^{-k^2\pi^2 t/2} \cos(k\pi d)\ , \tag{2}$$

where n is a positive large interger and a_k is

$$a_k = \begin{cases} 1 & k = 0 \\ \frac{1}{n}\sum_{i=1}^{n} cos(k\pi d_i\ ) & k = 1, 2, ..., n-1 \end{cases}$$

for more detail of density estimation via heat diffusion see [27, 29, 28]. The Eq.2 is a alternative and adaptive form of

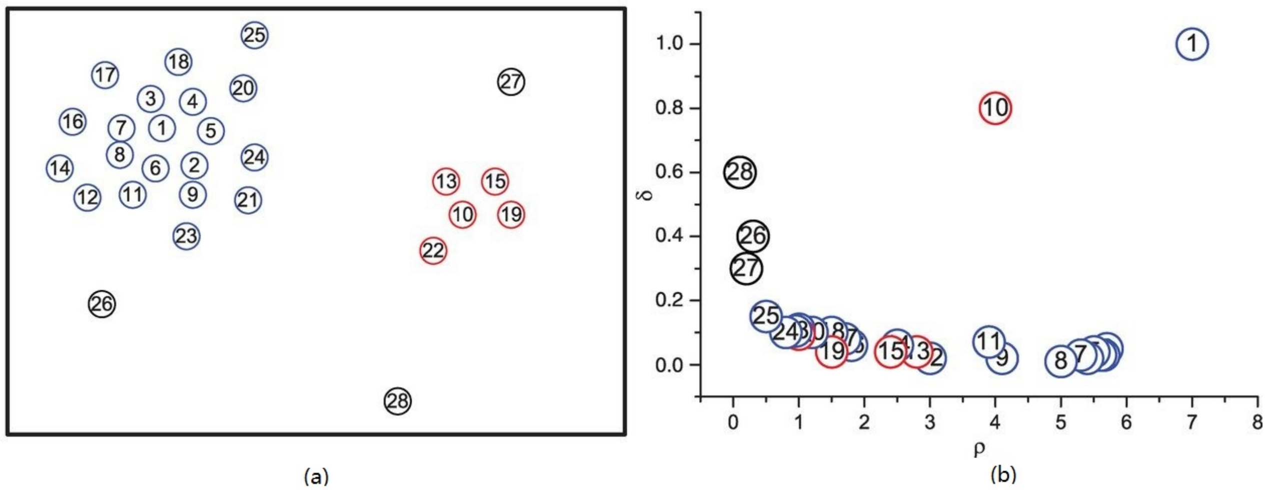

Figure 1: Decision graph representation of DP-clustering [24]

kernel density estimation (KDE) and also accounts both for optimal band width selection and boundary correction. The computational complexity of Eq.2 is $\mathcal{O}(n \log_2 n)$[27, 29] by utilizing fast Fourier transform. The Algorithm 4 is used to estimate the density of underlying datasets, adaptively.

Algorithm 4 Local density estimation

Require: D, distance matrix
Ensure: ρ, n length density vector
  $\rho \leftarrow$ esimated using Eq.2.

Unlike the density estimation methods introduced in DP, our proposed method is adaptive in nature to select an appropriate cutoff distance and is also capable of expressing true densities at border regions effectively.

Distance from higher local density points: In the proposed method, Algorithm 2 is used to measure the distance from the nearest higher density point. δ_i is the distance between point i and the nearest higher density point.

Pruning of non-cluster center points: In DP, the selection of cluster centers is a human intervention process that utilizes a heuristic approach called a decision graph. The decision graph is a main problem to automatic selection of cluster centers. However, in the proposed method, we introduce a pruning technique to remove non-cluster points and noise from cluster centers in two steps.

Pruning technique to detect local densities: According to DP, cluster centers are points that have high density and large distance compared to non-cluster centers. By applying this rule, all local dense points can be efficiently separated from cluster points. To detect all local dense points, we take the standard deviation value of δ and then subtract this value from the vector of distance (δ). Algorithm 5 is proposed to separate expected cluster centers and noise from non-cluster center points.

In Algorithm 5, parameter t is the time of the heat equation and β is the scaling parameter; its default value is 2. The time complexity to find the index of the expected density centers is $\mathcal{O}(n)$. In the next step, we remove the noise from the expected density centers. In this step, the noise is separated by utilizing the definition of noise given in DP, the noise is characterized as high values of δ with very low values of ρ, to remove noise, we take the mean of ρ vector and then subtract $\rho/1.5$ value from all Exp_C points. After successfully removing noise from the expected cluster centers, actual cluster centers are obtained.

Algorithm 5 Index of Expected Density Centers

Require: δ,$NNdistance$ vector of n objects
Ensure: Exp_C, indexes of expected density centers
  for $i \leftarrow 1 : n$ **do**
    $\delta(i) \leftarrow \delta(i) - \pi * \beta * t$
    if $\delta(i) >= \sigma(\delta)$ **then**
      $Exp_C(i) \leftarrow i$
    end if
  end for

Algorithm 6 Separation of noise from expected cluster centers

---

Require: ρ, n length density vector ; Exp_C, expected cluster centers
Ensure: C, indexes of cluster centers
  for $i \leftarrow 1 : Size(Exp_C)$ **do**
    if $\rho(Exp_C(i)) > mean(\rho)/1.5$ **then**
      $C(i) \leftarrow Exp_C(i))$
    end if
  end for

The Algorithm 6 is used to filter noise from cluster centers. The time complexity depends on the number of expected density centers and detected noise. From experiments, it is shown that the noise and cluster centers are always much less numerous than cluster points. After successfully identification of cluster centers Algorithm 3 is used to assigned labels to cluster centers.

4. EXPERIMENTS

The robustness of the proposed method is evaluated on 15 synthetic and real world datasets, Table 1 shows the detailed description about the datasets such as name, data points in dataset, dimension of the dataset and the source

Table 1: Description of tested datasets

Dataset	Classes	Objects	Dimensions	Source
Point Distibutions	4	2000	2	[24]
Wine	3	178	13	[30]
Aggregation	7	788	2	[31]
flame	2	240	2	[32]
Concave	2	2730	2	[33]
Path-based spiral	2	312	2	[34]
R15	15	600	2	[35]
Two Diamond	2	800	2	[36]
D31	31	3100	2	[35]
Dim2	9	1650	2	[37]
Toys problem	3	300	2	[38]
A1	20	3000	2	[39]
Diamond	9	3000	2	[40]
S1	15	5000	2	[41]
Leukemia	3	38	999	[42]

of dataset. The Rand indexing is used to measure the similarity of clusters created by the proposed method and DP, AP, k-medoids, and k-means.

To analyze the performance of the proposed Adaptive-DP for small-size datasets, benchmark datasets such as flame, toys problem, path-base spiral, and wine datasets are used. In the flame dataset, the Adaptive-DP successfully organized clusters into two clusters. Compared to DP, the proposed method automatically detected the cluster centers efficiently. In the flame dataset, the actual cluster centers are very dense compared to non-cluster center density. Hence, expected cluster centers are separated by utilizing Algorithm 5, and Algorithm 6 is utilized to differentiate between the expected noise points and actual cluster centers. Figure 2(a) shows the graphical representation of Adaptive-DP to separate cluster centers from non-cluster centers. Just for better understanding and explanation of the proposed method, noise and non-center points are set to zero distance to separate them from actual cluster centers. So simply detect the non-zero distance points as cluster centers. Figure 2(b) shows the final clusters created by the proposed method. As with the flame dataset, the Adaptive-DP also successfully organized the toys problem dataset into two clusters. Figure 2(c) shows the graphical representation of cluster center and non-cluster center points. In Figure 2(c) the distance of non-cluster centers is adjusted to zero and cluster centers are shown as high distance points. Figure 2(d) shows the separated clusters of the toys problem dataset created by the proposed Adaptive-DP method. In the path-based spiral dataset, Adaptive-DP successfully discovered the density connected points and discovered three clusters as shown in Figure 2(e). We utilized the wine dataset to benchmark the Adaptive-DP method on multi-dimensional datasets. The clusters that Adaptive-DP organized for the wine dataset are shown in Figure 2(f). All of the experimental results of the aforementioned datasets validate the robustness of the Adaptive-DP method on small and multi-dimensional datasets.

The aggregation synthetic dataset is utilized to evaluate the performance of Adaptive-DP to merge local densities into a single cluster. On the aggregation dataset, some clusters consist of more than one local density peak. To merge the peaks, only one scaling parameter (β) is used in Algorithm 5. The organized clusters can be refined by scaling β. At β =2 , the number of clusters detected by Adaptive-DP is presented in Figure 3(a). In Figure 3(b) the organized clusters of aggregation datasets are shown. However, the numbers of clusters detected by scaling the value of are shown in Figure 3(c). By default, the value of is adjusted to 2; however, it can be scaled to increase the number of clusters or to decrease the number of desired clusters. To obtain the minimum number of clusters, scaling toward a higher value is suggested, and to obtain more clusters, it is suggested to decrease the value of . Figure 3(d) shows the exact final organized clustering by the Adaptive-DP clustering method.

The performance of Adaptive-DP on the point distribution, diamond, and Dim-2 datasets was also evaluated. The point distribution consisted of noise with 5 clusters. Hence, it is proved that the Adaptive-DP method is also effective at filtering noise from cluster center points. Figure 4(a) shows the point distribution clusters organized by Adaptive-DP. We utilize the diamond dataset to evaluate the method's ability to effectively separate the highly connected edges of the clusters, as shown in Figure 4(b). The organized clusters of Dim-2 created by Adaptive-DP are shown in Figure 4(c).

To benchmark the Adaptive-DP method on large datasets, we utilized the S1, A1, D31, and concave datasets. In S1, A1, and D31, each cluster consisted of a single density peak; however, in Concave, different density peaks constituted a single cluster. In both single density cluster and multi-density peak clusters, the Adaptive-DP successfully identified the exact number of clusters. These large-size datasets are evidence that the Adaptive-DP method is equally effective for small- and large-size datasets.

In addition, the effectiveness of Adaptive-DP on different datasets (two-diamonds, R15) was evaluated. In the two-diamond dataset, the proposed method successfully identified two clusters; in the R15 dataset, it found 15 clusters datasets.

At last, we applied the approach on leukemia cancer, dataset, to identify three distinct subtypes, AML, ALL, and B-lineage ALL. The dataset possess serious challenges to various clustering approaches because of the high dimensions with small number of observations. Our approach successfully identify distinct three cancer subtypes and organized into groups as shown in Fig.5 (a). The visual representation of estimated density are shown in Fig.5(b), which reveals three distinct density regions exist in dataset. Most of clustering approach could not find the distinct samples, exactly. The Rand index score of AP, Hierarchical, Spectral, Density Peaks, and Adaptive-DP clustering are shown in Fig. 5(c).

4.1 Comparisons

To validate the performance of the Adaptive-DP method, we conducted comprehensive comparison with state-of-the-art methods. We used the Rand Index to measure the accuracy of formulated clusters. We evaluated and compared the Rand Index of Adaptive-DP to that of other famous clustering methods on four different synthetic clustering datasets.

Rand measuring or Rand index in the field of clustering is to measure the similarity between two data clusterings. Actually, it is the ratio of data correctly clustered out of all possible pairs. The Rand index uses a pairwise approach to evaluate True Negatives (TN), True Positives (TP), False

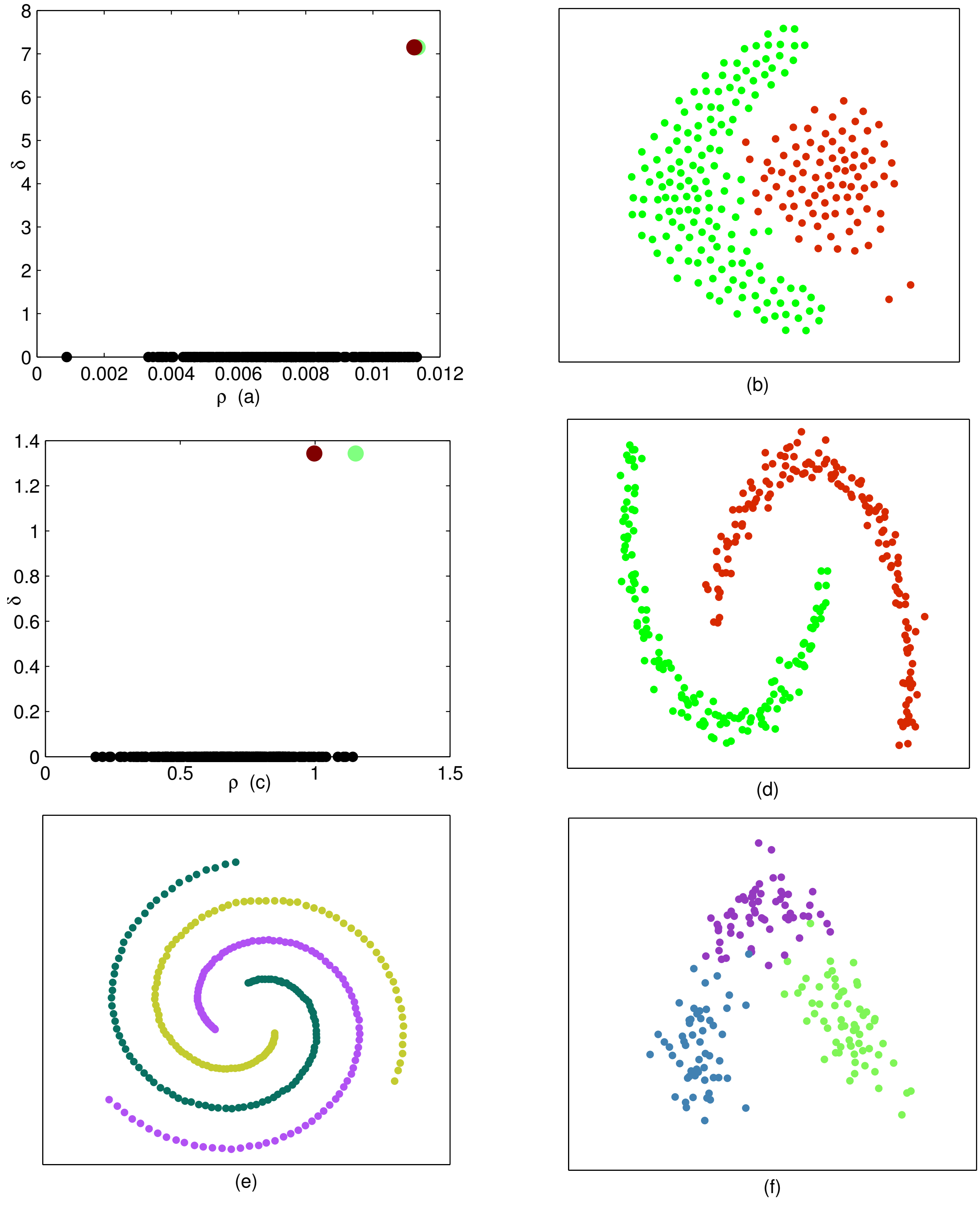

Figure 2: Evaluation of Adaptive-DP method on small multi-dimensional datasets.(a) Detection of exact number of clusters without utilizing the decision graph,large distance points are cluster centers of flame dataset.(b) flame clusters organized by proposed method.(c,d) The detection of exact cluster centers and clusters organized by Adaptive-DP method of toys problem dataset, respectively. (e) Organized clusters of path-based spiral dataset by using Adaptive-DP method.(f) clusters of wine dataset created by proposed Adaptive-DP method.

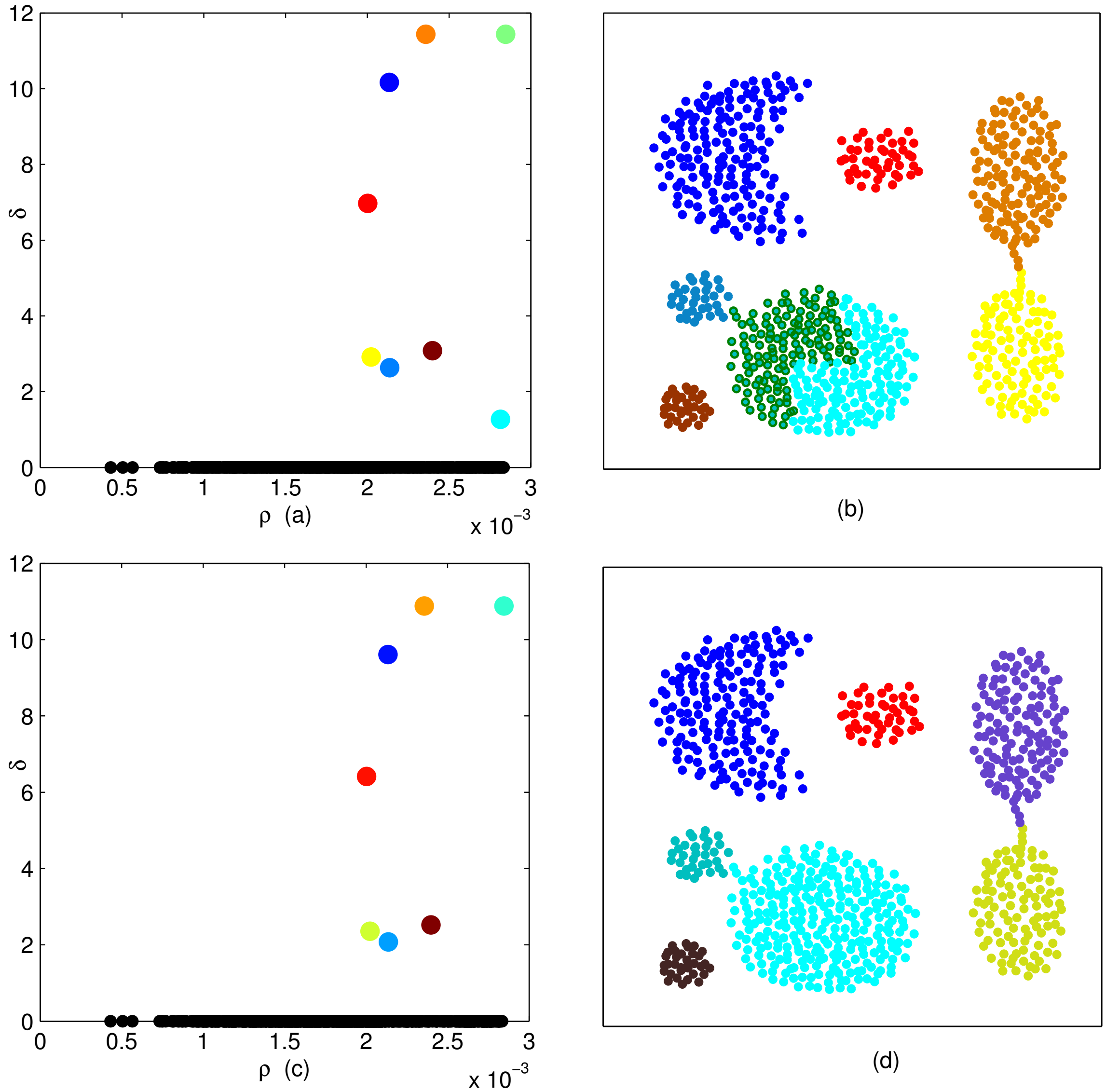

Figure 3: Organized clusters of aggregation dataset by proposed method at different values of β. (a) Eight cluster centers are detected at the β=2 and in (b) the seven clusters are shown. (c) The organized clusters are minimize by increasing the value β=4 and seven perfect clusters are shown in (d).

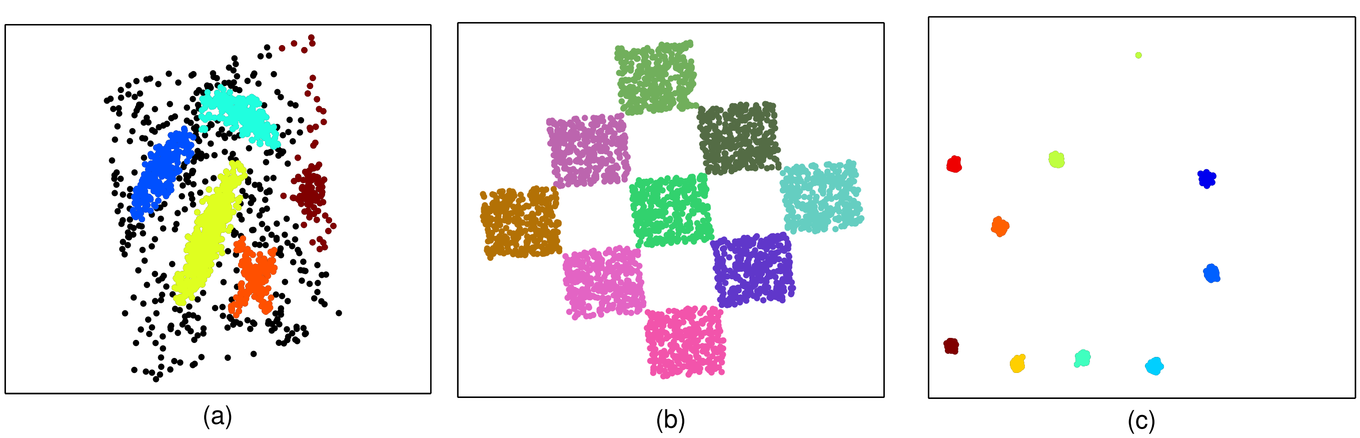

Figure 4: Adaptive-DP organized clusters of (a) point distribution, (b) diamond, and (c) Dim-2 datasets.

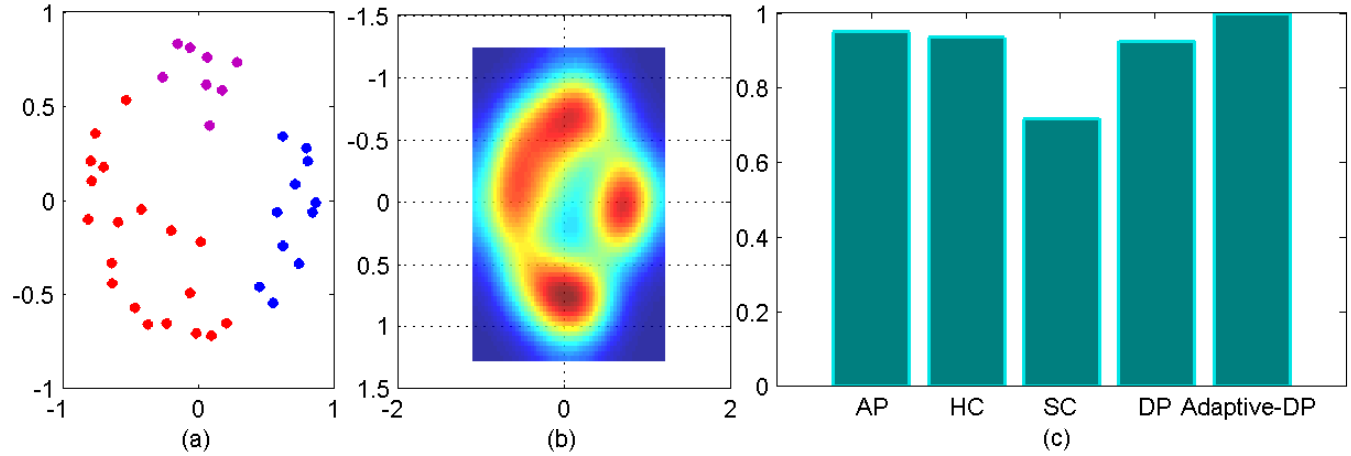

Figure 5: Analysis and comparison of clustering algorithms on leukemia dataset.(a) Resultant clusters separated by proposed method, we successfully identified and group dataset into three clusters consisted of 11 AML, 8 ALL, and 19 B-lineage ALL samples. (b) Visual representation of density estimated by proposed method. (c) The rand index based comparison of our approach with AP, Hierarchal, Spectral, and DP clustering methods.

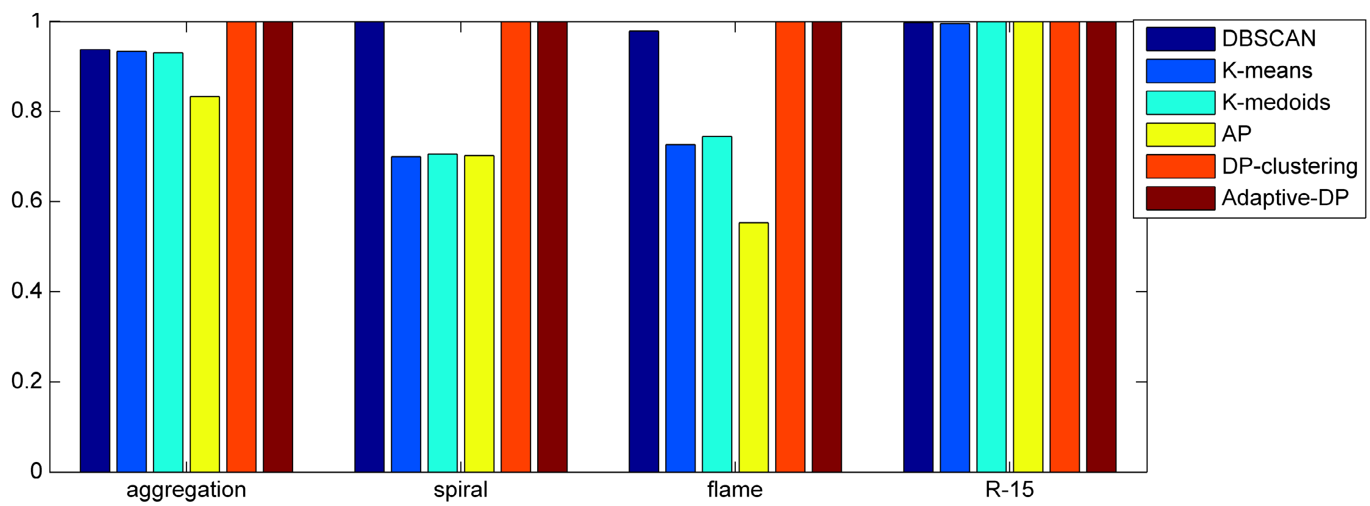

Figure 6: Comparison of Rand index of proposed method with state of the art and famous clustering methods, over 4 synthetic clustering datasets.

Negatives (FN), and False Positives (FP. The Rand index can be measured utilizing the following expression[42]:

$$R = \frac{TP + TN}{TP + FP + FN + TN}$$

The comparison of the Rand index on four datasets with dbscan, k-means, k-medoids, AP, DP-Clustering, and Adaptive-DP is shown in Figure 6. Figure 6 expresses the effectiveness of the given method compared with state of the art clustering methods.

5. CONCLUSION

In DP-clustering, the decision graph based approach is used to manually select the exact number of clusters. In this paper, we have presented a new method (Adaptive-DP) that estimates the density by using a heat diffusion method and an adaptive approach to select the exact number of cluster centers. The limitations of DP-the difficulty of selecting an appropriate method to estimate density, selection of cutoff distance, and the human interpretation required to select the number of cluster centers-are improved in Adaptive-DP. The experimental results on 14 datasets and comparison with state-of-the-art methods show the robustness and effectiveness of the proposed method.

6. REFERENCES

[1] A. Bouroumi, M. Limouri, A. Essa, et al. Unsupervised fuzzy learning and cluster seeking. *Intelligent Data Analysis*, 4(3, 4):241–253, 2000.

[2] R. Bie, R. Mehmood, S. Ruan, Y. Sun, and H. Dawood. Adaptive fuzzy clustering by fast search and find of density peaks. *Personal and Ubiquitous Computing*, 20(5):785–793, 2016.

[3] M. Shoaib, A. Daud, and M. Khiyal. Improving similarity measures for publications with special focus on author name disambiguation. *Arabian Journal for Science & Engineering (Springer Science & Business Media BV)*, 40(6), 2015.

[4] J. Ramon. Clustering and instance based learning in first order logic. *AI Communications*, 15(4):217–218, 2002.

[5] R. Mehmood, R. Bie, L. Jiao, H. Dawood, and Y. Sun. Adaptive cutoff distance: Clustering by fast search and find of density peaks. *Journal of Intelligent & Fuzzy Systems*, 31(5):2619–2628, 2016.

[6] Y. Li, S. Feng, X. Zhang, and L. Jiao. Sar image segmentation based on quantum-inspired multiobjective evolutionary clustering algorithm. *Information Processing Letters*, 114(6):287–293, 2014.

[7] R. Weng, J. Lu, J. Hu, G. Yang, and Y.-P. Tan. Robust feature set matching for partial face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 601–608, 2013.

[8] J. Yu, R. Hong, M. Wang, and J. You. Image clustering based on sparse patch alignment framework. *Pattern Recognition*, 47(11):3512–3519, 2014.

[9] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. A survey on cyber security for smart grid communications. *IEEE Communications Surveys and tutorials*, 14(4):998–1010, 2012.

[10] H. U. Khan, A. Daud, U. Ishfaq, T. Amjad, N. Aljohani, R. A. Abbasi, and J. S. Alowibdi. Modelling to identify influential bloggers in the blogosphere: A survey. *Computers in Human Behavior*, 68:64–82, 2017.

[11] A. Daud, M. Ahmad, M. Malik, and D. Che. Using machine learning techniques for rising star prediction in co-author network. *Scientometrics*, 102(2):1687–1711, 2015.

[12] K. M. Jones, M. Lacy, et al. Measuring the clustering around normal and dust-obscured quasars at 2 in the spitzer extragalactic representative volume survey (servs). In *American Astronomical Society Meeting Abstracts# 223*, volume 223, 2014.

[13] T. Lefèvre, C. Rondet, I. Parizot, and P. Chauvin. Applying multivariate clustering techniques to health data: the 4 types of healthcare utilization in the paris metropolitan area. *PloS one*, 9(12):e115064, 2014.

[14] C. Xu and Z. Su. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*, page btv088, 2015.

[15] D. Jaeger, J. Barth, A. Niehues, and C. Fufezan. pygcluster, a novel hierarchical clustering approach. *Bioinformatics*, 30(6):896–898, 2014.

[16] P. Tellaroli, M. Bazzi, M. Donato, A. R. Brazzale, and S. Drăghici. Cross-clustering: A partial clustering algorithm with automatic estimation of the number of clusters. *PloS one*, 11(3):e0152333, 2016.

[17] M. J. Paszek, D. Boettiger, V. M. Weaver, and D. A. Hammer. Integrin clustering is driven by mechanical resistance from the glycocalyx and the substrate. *PLoS Comput Biol*, 5(12):e1000604, 2009.

[18] H. Fushing, H. Wang, K. VanderWaal, B. McCowan, and P. Koehl. Multi-scale clustering by building a robust and self correcting ultrametric topology on data points. *PloS one*, 8(2):e56259, 2013.

[19] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

[20] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[21] G. H. Shah, C. Bhensdadia, and A. P. Ganatra. An empirical evaluation of density-based clustering techniques. *International Journal of Soft Computing and Engineering (IJSCE) ISSN*, 22312307:216–223, 2012.

[22] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

[23] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.

[24] A. Rodriguez and A. Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

[25] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.

[26] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8):790–799, 1995.

[27] S. Krishnaswamy, M. H. Spitzer, M. Mingueneau, S. C. Bendall, O. Litvin, E. Stone, D. PeŠer, and G. P. Nolan. Conditional density-based analysis of t cell signaling in single-cell data. *Science*, 346(6213):1250689, 2014.

[28] R. Mehmood, G. Zhang, R. Bie, H. Dawood, and H. Ahmad. Clustering by fast search and find of density peaks via heat diffusion. *Neurocomputing*, 208:210–217, 2016.

[29] Z. I. Botev, J. F. Grotowski, D. P. Kroese, et al. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010.

[30] M. Forina et al. An extendible package for data exploration, classification and correlation. *Institute of Pharmaceutical and Food Analisys and Technologies, Via Brigata Salerno*, 16147, 1991.

[31] A. Gionis, H. Mannila, and P. Tsaparas. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):4, 2007.

[32] L. Fu and E. Medico. Flame, a novel fuzzy clustering method for the analysis of dna microarray data. *BMC bioinformatics*, 8(1):3, 2007.

[33] Y. Ge and S. C. Sealfon. flowpeaks: a fast unsupervised clustering for flow cytometry data via k-means and density peak finding. *Bioinformatics*, 28(15):2052–2058, 2012.

[34] H. Chang and D.-Y. Yeung. Robust path-based spectral clustering. *Pattern Recognition*, 41(1):191–203, 2008.

[35] C. J. Veenman, M. J. T. Reinders, and E. Backer. A maximum variance cluster algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1273–1280, 2002.

[36] M. Khalilia and M. Popescu. Topology preservation in fuzzy self-organizing maps. In *WCSC*, pages 105–114. Springer, 2013.

[37] P. Franti, O. Virmajoki, and V. Hautamaki. Fast agglomerative clustering using a k-nearest neighbor graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1875–1881, 2006.

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[39] I. Karkkainen and P. Franti. Dynamic local search for clustering with unknown number of clusters. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 240–243. IEEE, 2002.

[40] S. Salvador and P. Chan. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on*, pages 576–584. IEEE, 2004.

[41] P. Fränti and O. Virmajoki. Iterative shrinking method for clustering problems. *Pattern Recognition*, 39(5):761–775, 2006.

[42] C. Wiwie, J. Baumbach, and R. Röttger. Comparing the performance of biomedical clustering methods. *Nature methods*, 12(11):1033–1038, 2015.