Clustering with or without the Approximation*

Frans Schalekamp[1], Michael Yu[2], and Anke van Zuylen[1]

[1] ITCS, Tsinghua University
{frans,anke}@tsingua.edu.cn
[2] MIT
mikeyu@mit.edu

Abstract. We study algorithms for clustering data that were recently proposed by Balcan, Blum and Gupta in SODA'09 [4] and that have already given rise to two follow-up papers. The input for the clustering problem consists of points in a metric space and a number k, specifying the desired number of clusters. The algorithms find a clustering that is provably close to a target clustering, provided that the instance has the "$(1+\alpha,\varepsilon)$-property", which means that the instance is such that all solutions to the k-median problem for which the objective value is at most $(1+\alpha)$ times the optimal objective value correspond to clusterings that misclassify at most an ε fraction of the points with respect to the target clustering. We investigate the theoretical and practical implications of their results.

Our main contributions are as follows. First, we show that instances that have the $(1+\alpha,\varepsilon)$-property and for which, additionally, the clusters in the target clustering are large, are easier than general instances: the algorithm proposed in [4] is a constant factor approximation algorithm with an approximation guarantee that is better than the known hardness of approximation for general instances. Further, we show that it is NP-hard to check if an instance satisfies the $(1+\alpha,\varepsilon)$-property for a given (α,ε); the algorithms in [4] need such α and ε as input parameters, however. We propose ways to use their algorithms even if we do not know values of α and ε for which the assumption holds. Finally, we implement these methods and other popular methods, and test them on real world data sets. We find that on these data sets there are no α and ε so that the dataset has both $(1+\alpha,\varepsilon)$-property and sufficiently large clusters in the target solution. For the general case, we show that on our data sets the performance guarantee proved by [4] is meaningless for the values of α,ε such that the data set has the $(1+\alpha,\varepsilon)$-property. The algorithm nonetheless gives reasonable results, although it is outperformed by other methods.

* This work was supported in part by the National Natural Science Foundation of China Grant 60553001, and the National Basic Research Program of China Grant 2007CB807900, 2007CB807901. Part of this work was done while the second author was visiting the Institute for Theoretical Computer Science at Tsinghua University in the summer of 2009.

M.T. Thai and S. Sahni (Eds.): COCOON 2010, LNCS 6196, pp. 70–79, 2010.

1 Introduction

Clustering is an important problem which has applications in many situations where we try to make sense of large amounts of data, such as in biology, marketing, information retrieval, et cetera. A common approach is to infer a distance function on the data points based on the observations, and then to try to find the correct clustering by solving an optimization problem such as the k-median problem, k-means problem or min-sum clustering problem. Unfortunately, these three optimization problems are all NP-hard, hence we do not expect to find algorithms that find the optimal solution in polynomial time. Research has therefore focused on finding good heuristics (such as for example the popular k-means++ algorithm [1]), exact methods (see for example [8]), and approximation algorithms: polynomial time algorithms that come with a guarantee β that the returned solution has objective value at most β times the optimum value. Research into approximation algorithms for these three clustering problems has produced a large number of papers that demonstrate approximation algorithms as well as lower bounds on the best possible guarantee. However, in many cases there is still a gap between the best known approximation algorithm and the best known lower bound.

In a recent paper, Balcan, Blum and Vempala [5] (see also Balcan, Blum and Gupta [4]) observe the following: The optimization problems that we try to solve are just proxies for the real problem, namely, finding the "right" clustering of the data. Hence, if researchers try so hard to find better approximation algorithms, that must mean that we believe that this will help us find clusterings that are closer to the target clustering. More precisely, Balcan, Blum and Vempala [5] turn this implicit belief into the following explicit assumption: there exist $\alpha > 0, \varepsilon > 0$ such that any solution with objective value at most $(1+\alpha)$ times the optimum value misclassifies at most an ε fraction of the points (with respect to the unknown target clustering). We will call this the $(1+\alpha, \varepsilon)$-property. By making this implicit assumption explicit, Balcan et al. [4] are able to show that, given (α, ε) such that the $(1+\alpha, \varepsilon)$-property holds, quite simple algorithms will give a clustering that misclassifies at most an $O(\varepsilon)$-fraction of the points. In the case when the clusters in the target clustering are "large" (where the required size is a function of ε/α), they give an algorithm that misclassifies at most an ε fraction of the points. In the general case, they give an algorithm that misclassifies at most an $O(\varepsilon/\alpha)$-fraction. They do not need better approximation algorithms for the k-median problem to achieve these results: in fact, [4] shows that finding a $(1+\alpha)$-approximation algorithm does not become easier if the instance satisfies the $(1+\alpha, \varepsilon)$-property.

These results seem quite exciting, because they allow us to approximate the target clustering without approximating the objective value of the corresponding optimization problem. As Balcan et al. [4] point out, especially if approximating the objective to within the desired accuracy is hard, we have no choice but to "bypass" the objective value if we want to approximate the target clustering.

However, it is not immediately clear how useful these results are in practice. A first concern is that the algorithms need parameters α and ε such that the instance satisfies the $(1+\alpha, \varepsilon)$-property. The paper by Balcan et al. [4] gives no suggestions on how a practitioner can find such α and ε. And of course an interesting question is whether these new algorithms outperform previously known methods in approximating the target clustering, if we do know α and ε, especially in the case when α is smaller than the guarantee of the best known approximation algorithm.

In this paper, we set out to investigate practical and theoretical implications of the algorithms in Balcan et al. [4] We now briefly describe our contributions.

1.1 Our Contributions

We focus on the case when the optimization problem we need to solve is the k-median problem. Our main theoretical contribution is a proof that the algorithm for "large clusters" given by Balcan et al. [4] is in fact an approximation algorithm with a guarantee $1 + \frac{1}{1/2+5/\alpha}$. One could argue that the algorithm of Balcan et al. [4] is most interesting when $\alpha \leq 2$ (since for $\alpha > 2$ one can use the algorithm of Arya et al. [2] to obtain the claimed result), and hence in those cases their algorithm has an approximation guarantee of at most $\frac{4}{3}$. However, for the general case of the k-median problem, there is a hardness of approximation of $1 + \frac{1}{e}$ [11]. As $1 + \frac{1}{e} \approx 1.37$ is larger than $\frac{4}{3}$, this means that these instances are provably easier than the general class of instances. We note that Balcan et al. [4] show that approximating the k-median objective does *not* become easier if we are guaranteed that the instance has the $(1+\alpha, \varepsilon)$-property. We show that it *does* become easier for those instances that have the $(1+\alpha, \varepsilon)$-property and for which the clusters in the target clustering are "large".

For the general case, we show that it is NP-hard to check whether a data sets satisfies the $(1+\alpha, \varepsilon)$-property for a given α, ε; however, knowledge of such parameters is necessary to run the algorithm of Balcan et al. [4].

We implement the algorithms of Balcan et al. [4] and compare the results to the outcome of previously known methods for various real world data sets. We show how to efficiently run the algorithms for all possible values of the parameters (α, ε) (regardless of whether the assumption holds for the pair of values), and suggest a heuristic for choosing a good solution among the generated clusterings. The algorithm for "large clusters" fails to find a solution on any of our instances and for any value of the parameters α, ε. The algorithm for the general case, however, does return reasonably good solutions. We compare these results to other methods, and find that they are reasonable, but that there are other methods, both heuristics and approximation algorithms, which are significantly better both in terms of approximating the target clustering and approximating the k-median objective.

We also show how to enumerate all values of α, ε for which the $(1+\alpha, \varepsilon)$-property holds, which we note are not practical as they need or calculate the optimal k-median solution. We find that indeed our data sets never satisfy the $(1+\alpha, \varepsilon)$-property *and* the large clusters assumption. For the general case, we find that the proven guarantee on the misclassification of $O(\varepsilon/\alpha)$ is greater than one.

1.2 Related Work

Due to space constraints, we focus our discussion on research that is similar in spirit to our work, in the sense that it restricts the input space and uses these additional assumptions to obtain improved algorithmic results.

Balcan, Blum and Vempala [5] study the problem of approximating an unknown target clustering given a distance metric on the points. They identify properties of the distance metric that allow us to approximate an unknown target clustering. One of the properties they define is the $(1+\alpha, \varepsilon)$-property which is exploited by Balcan, Blum and Gupta [4] to find clusterings that are provably close to the target clustering. We describe their work in more detail in the next section. Two follow-up papers have extended their results in two directions: Balcan, Röglin and Teng [7] consider the setting where all but a γ-fraction of the points have the $(1+\alpha, \varepsilon)$-property and this γ-fraction of points is adversarially chosen. Balcan and Braverman [6] improve the results in [4] if the goal is to approximate the target clustering and the input satisfies the $(1+\alpha, \varepsilon)$-property with respect tot the min-sum objective.

Ostrovsky, Rabani, Schulman and Swamy [15] identify natural properties under which variants of Lloyd's algorithm are guaranteed to quickly find near-optimal solutions to the k-means problem. The recent paper of Bilu and Linial [10] gives polynomial time algorithms for a certain class of inputs to the max-cut problem, which they call "stable" instances. There are strong similarities between [4] and [10]: both approaches define classes of inputs, for which they can give algorithms that perform better than what is possible for general instances. In fact, it is possible to show that the $(1+\alpha, \varepsilon)$-property implies a stability property in similar vein to the stability property defined in [10]. Theorem 1 in this paper shows that the class of inputs defined by $(1+\alpha, \varepsilon)$-property combined with the assumption that the clusters of the target clustering are large, is easier to approximate than general instances of the k-median problem.

2 Problem Definition

2.1 k-Median Problem

In the k-median problem, we are given a set of elements X and a distance function $d : X \times X \to \mathbb{R}^{\geq 0}$ which forms a metric (i.e., d satisfies the triangle inequality), a subset of elements $V \subset X$ that need to be covered and a parameter $k \in \mathbb{N}$. We denote $|V| = n$. The goal is to choose k cluster centers $v_1, \ldots, v_k \in X$ so as to minimize $\sum_{u \in V} \min_{i=1,\ldots,k} d(u, v_i)$.

We denote by OPT the optimum ojective value of a given instance, and we say an algorithm is an β-approximation algorithm for the k-median problem if for any instance it is guaranteed to output cluster centers $v_1, \ldots, v_k \in X$ so that $\sum_{u \in V} \min_{i=1,\ldots,k} d(u, v_i) \leq \beta OPT$.

2.2 Setting of Balcan, Blum and Gupta

In the setting considered by Balcan et al. [4], an instance also includes an unknown target clustering, i.e. a partition $C_1^*, \ldots, C_k^*$ of V. We say a clustering $C_1, \ldots, C_k$ is ε-close to the target clustering, if there exists a permutation π such that $\frac{1}{n}\sum_{i=1}^{k} |C_i^* \backslash C_{\pi(i)}| \leq \varepsilon$. The misclassification is defined as the smallest ε such that the clustering is ε-close to the target clustering.

When clustering data into k clusters, an often used approach is to define a distance function on the data based on observations, and to solve an optimization problem (for example, the k-median problem) to obtain a clustering. Balcan et al. [4] argue that the quest for better approximation algorithms thus implies a belief that better approximations will result in solutions that are closer to the unknown target clustering. They formalize this implicit belief into the following property.

Definition 1 ($(1+\alpha,\varepsilon)$-property). *An instance satisfies the $(1+\alpha,\varepsilon)$-property, if any k-median solution with objective value at most $(1+\alpha)OPT$ is ε-close to the target clustering.*

Balcan et al. [4] propose and analyze two algorithms: the first one is for instances that both satisfy the $(1+\alpha,\varepsilon)$-property and additionally are such that the clusters in the target clustering are of size at least $(3+10/\alpha)\varepsilon n+2$. This algorithm needs α, ε and OPT as inputs. (There is a way to get around having to give OPT as an input, as is shown in [4].) The algorithm is guaranteed to return a clustering that is ε-close to the target clustering.

The second algorithm is for instances that (just) satisfy the $(1+\alpha,\varepsilon)$-property. This algorithm is remarkably simple, but also assumes knowledge of α, ε and OPT. This algorithm for the less restricted input space is guaranteed to return a clustering that is $O(\varepsilon/\alpha)$-close to the target clustering. Instead of OPT an approximate value can be used, at the cost of a deteriorating guarantee: If the value is guaranteed to be at most βOPT, then the misclassification guarantee becomes $O(\varepsilon\beta/\alpha)$. In the remainder of this paper, we will use the abbreviations BBGlarge and BBGgeneral to refer to these algorithms.

Instead of working with the $(1+\alpha,\varepsilon)$-property directly, it is usually easier to work with a weaker property that is implied by the $(1+\alpha,\varepsilon)$-property (as shown in Lemma 3.1 in [4]). We will refer to this as the "weak $(1+\alpha,\varepsilon)$-property". Note that this weaker property does not depend on a target clustering.

Definition 2 (weak $(1+\alpha,\varepsilon)$-property if all clusters in the target clustering have more than $2\varepsilon n$ points). *In the optimal k-median solution, there are at most εn points for which the second closest center is strictly less than $\alpha OPT/(\varepsilon n)$ farther than the closest center.*

Definition 3 (weak $(1+\alpha,\varepsilon)$-property). *In the optimal k-median solution, there are at most $6\varepsilon n$ points for which the second closest center is strictly less than $\alpha OPT/(2\varepsilon n)$ farther than the closest center.*

3 Theoretical Aspects of the BBG Algorithms

We now show that in fact the algorithm BBGlarge succeeds in finding a solution which not only has a classification error of at most ε, but is also an approximately optimal k-median solution.

Theorem 1. *If the k-median instance satisfies the $(1+\alpha,\varepsilon)$-property and each cluster in the target clustering has size at least $(3+10/\alpha)\varepsilon n+2$, then the algorithm for large clusters proposed by Balcan et al. [4] gives a $(1+1/(1/2+5/\alpha))$-approximation algorithm for k-median clustering.*

Proof (sketch). In the proof of the fact that the algorithm for large clusters only misclassifies an ϵ fraction of the points, Balcan et al. [4] distinguish two types of points; red points, and non-red (green or yellow) points. They show that the only points that are potentially misclassified are the red points, and there are only ϵn red points.

To show the approximation result, we first repeat the final step of their algorithm: Given clusters $C_1,\ldots,C_k$, we compute for each point x the index $j(x)$ of the cluster that minimizes the median distance from x. The new clustering is obtained by letting $C'_j=\{x : j(x)=j\}$. It follows from the analysis of Balcan et al. [4] that the new clustering again only misclassifies red points. Moreover, we know that for each misclassified red point x, (say in cluster i instead of cluster j), there is a large number of "good reasons" why we made this mistake: pairs of non-red points $y(i),y(j)$ where $y(i)$ is in cluster i and $y(j)$ is in cluster j, and x is closer to $y(i)$ than to $y(j)$. In order to prove an approximation guarantee, we need to bound (or "charge") the difference between the distance from x to its cluster center in our solution and in the optimal solution. By using the triangle inequality, it is not hard to show that we can charge this difference against the sum of the distances from $y(i)$ and $y(j)$ to their respective cluster centers. Since $y(i)$ and $y(j)$ are non-red, and hence correctly clustered, we charge against a piece of the optimal solution. And, since there are "many" pairs $y(i),y(j)$, we only need to charge each piece a "small" number of times, where small turns out to be $1/(1/2+5/\alpha)$. We refer the reader to the full version [16] for further details. □

We note that, although we need to know values of α,ε to get the guaranteed bound on misclassification, we do not need it to obtain the approximation result in Theorem 1: It is not hard to show that we can try all relevant values of α and ε in polynomial time, and get the approximation result by returning the solution with smallest objective value.

We finally remark that the algorithms proposed by Balcan et al. [4] are most interesting for instances with a $(1+\alpha,\varepsilon)$-property with $\alpha\le 2$: if $\alpha>2$ then we could then find an ε-close clustering by running a $(3+2/p)$-approximation algorithm [2] for sufficiently large p. Hence for those α for which the Balcan et al. algorithms are interesting, we have shown that the large clusters algorithm gives a $\frac{4}{3}$-approximation algorithm.

Our second result in this section is to show that verifying if an instance has the $(1+\alpha, \varepsilon)$-property for a given α, ε is NP-hard. The proof is a reduction from max k-cover, and is given in the appendix of the full version of this paper [16].

Lemma 1. *It is NP-hard to verify whether an instance has the (weak)* $(1+\alpha, \varepsilon)$*-property for a given* α, ε*.*

We should remark two things about Lemma 1. First of all, in our proof we need to choose $\alpha \approx \varepsilon/n^3$. In that case, the guarantee given by Balcan et al. [4] is $O(\varepsilon/\alpha) = O(n^3)$, hence this does not constitute an interesting case for their algorithm. Second, our lemma does not say that it is NP-hard to find *some* α, ε for which the $(1+\alpha, \varepsilon)$-property holds. However, we do not know how to find such α, ε efficiently.

4 Practical Aspects of the BBG Algorithms

4.1 Data Sets

We use two popular sets of data to test the algorithms, and compare their outcomes to other methods. We use the pmed data sets from the OR-Library [9] to investigate whether the methods proposed by Balcan et al. [4] give improved performance on commonly used k-median data sets compared to known algorithms in either misclassification, objective value or performance relative to the running time. These instances are distance based but do not have a ground truth clustering. Note that the $(1+\alpha, \varepsilon)$-property implies that the optimal k-median clustering is ε-close to the target clustering (whatever the target clustering is), hence we can assume that the optimal k-median clustering is the target clustering while changing the misclassification of any solution with respect to the target by at most ε.

The second data sets we use come from the University of California, Irivine (UCI) Machine Learning Repository [3]. For these data sets a ground truth clustering is known and given. The data sets we use have only numeric attributes and no missing values. To get a distance functions, we first apply a "z-transform" on each of the dimensions, i.e., for each attribute we normalize the values to have mean 0 and standard deviation 1. Next, we calculate the Euclidean distance between each pair of points. We note that it may be possible to define distance functions that give better results in terms of approximating the target clustering. This is not within the scope of this paper, as we are only interested in comparing the performance of different algorithms for a given distance function.

4.2 Implementing the BBG Algorithms

Balcan et al. [4] do not discuss how to find values of α and ε to use. Indeed, Lemma 1 gives an indication that this may be far from trivial. It is not hard to realize however, that by varying α and ε, both the algorithms can generate only a polynomial number of different outcomes: In the case of BBGgeneral, α

and ε are used only to determine a threshold graph, for which there are only $O(n^2)$ possibilities. In the case of BBGlarge, a second graph is formed, which connects vertices that have at least b neighbors in the threshold graph, where b is a function of α and ε. This leads to a total of $O(n^3)$ possible outcomes for the BBGlarge algorithm. We therefore propose bypassing the fact that we do not know which values of α and ε to use by iterating over all possible outcomes. By being careful in the implementation, it is possible to get reasonably efficient algorithms; we refer the interested reader to the full version of this paper [16] for more details.

For the algorithm for the large clusters case, we found a somewhat surprising outcome: it did not return any clustering on any instance! This means that there exists no α, ε such that the data satisfied the $(1+\alpha, \varepsilon)$-property *and* the clusters in the target clustering had size at least $(3 + 10/\alpha)\varepsilon n + 2$.

For the algorithm for the general case, our implementation ideas reduce the number of solutions to consider from $O(n^2)$ to a much smaller number; at most 5% of the maximum possible number of solutions $n(n-1)/2$. Our next challenge is how to choose a good solution, that is close to the target clustering. A natural choice is to choose the outcome $C_1, \ldots, C_k$ with the lowest k-median objective value defined as $\sum_{i=1}^{k} \min_{c \in X} \sum_{v \in C_i} d(c, v)$. The following lemma shows that unfortunately the k-median objective is not always a reliable indicator of the best solution. The proof is given in the full version of this paper [16].

Lemma 2. *For a given instance, let δ be the misclassification of the solution with lowest k-median objective value among all outcomes obtained by the BBGgeneral algorithm for all threshold graphs. Then $\delta \neq O(\varepsilon/\alpha)$, even if the instance has the weak $(1 + \alpha, \varepsilon)$-property.*

We tried several other criteria for choosing a solution, which are inspired by the analysis of Balcan et al. [4]. None of these criteria is guaranteed to choose a solution with small misclassification, but some of them, including the k-median objective value, seem to work quite well in practice. Since the k-median objective value is quick to evaluate, we chose this criterion for our experimental comparison: on average, the misclassification of the best solution among all solutions generated is 6 percentage points lower than the misclassification of the solution with the best k-median objective value.

4.3 Verifying the $(1 + \alpha, \varepsilon)$-Property for These Data Sets

We found that all instances of our data sets do not lie in the restricted input space for which the BBG algorithm for large clusters is designed. The values of (α, ε) for which the weak $(1 + \alpha, \varepsilon)$-property for the general case (Definition 3) holds are such that the exact guarantee proved by Balcan et al., which is $(25 + 40/\alpha)\varepsilon$, is much larger than 1. However, ε/α was itself is always less than 1, so an $O(\varepsilon/\alpha)$ guarantee is meaningful for smaller constants. For more discussion we refer to the full version of this paper [16].

5 Comparison to Other Methods

We compare the quality and running time of the BBG algorithms to various heuristics and approximation algorithms for the k-median problem. More specifically, we implemented the following algorithms in MATLAB: the primal-dual algorithm proposed by Jain and Vazirani [13]; the primal-dual algorithm proposed by Jain, Mahdian, Markakis, Saberi and Vazirani [12]; Lloyd's algorithm [14]; k-means++ by Arthur and Vassilvitskii [1]; two variants of Local Search [2].

Due to space constraints, we refer the reader to the full version of this paper [16] for an overview of the outcome of the experiments. Although we found in the previous section that the theoretical guarantees of the BBGgeneral algorithm are meaningless for our data sets, it is clear from our experiments that the algorithm does give reasonable clusterings, and it is fast, even when checking all threshold graphs. However, for our data sets, other algorithms clearly outperform the BBGgeneral algorithm. In terms of overall performance, Local Search, which chooses random improving moves, is superior, both in terms of k-median objective and closeness to the target clustering. The algorithm of Jain et al. [12] is second in terms of performance, followed by kmeans++.

6 Conclusion and Open Problems

In this paper, we investigate theoretical and practical aspects of a new approach to clustering proposed by Balcan et al. [4]. We show that the assumption needed for their strongest result (the "large" clusters case) defines a set of "easy" instances: instances for which we can approximate the k-median objective to within a smaller ratio than for general instances. Our practical evaluations show that our instances do not fall into this category. For the algorithm for the general case, we give some theoretical justification that it may be hard to find the values of parameters α, ε that are needed as input. We show how to adapt the algorithm so we do not need to know these parameters, but this approach does not come with any guarantees on the misclassification. In our experimental comparison, the performance is reasonable but some existing methods are significantly better.

An interesting direction to evaluate the pratical performance of the algorithms by Balcan et al. [4] would be to test them on "easy" instances, i.e. instances for which the $(1+\alpha, \varepsilon)$-property holds for values ε, α for which ε/α is small, perhaps by identifying a small set of points whose removal ensures that this is the case, which was studied by Balcan, Röglin and Teng [7].

Theoretically, our results also raise the question whether it is possible to show an approximation guarantee for the algorithms for instances that satisfy the $(1+\alpha, \varepsilon)$-property and for which the target clustering has "large" clusters that were proposed for other objective functions, namely k-means and min-sum k-clustering, by Balcan et al. [4] and Balcan and Braverman [6]. For the general case, more research into exploiting this property may lead to algorithms which outperform existing methods. On the other hand, it would be interesting to have a lower bound on the misclassification of any (reasonable) algorithm when given an α, ε, such that the $(1+\alpha, \varepsilon)$-property holds. In particular, it would be interesting to know if the dependence on ε/α in either the guarantee on the misclassification or in the minimum cluster size is unavoidable.

Finally, an interesting direction is to find (other) classes of inputs defined by natural properties for which one can give algorithms that perform better than what is possible for the general class of inputs, both for the k-median problem and other optimization problems.

References

1. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: SODA '07: 18th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035 (2007)
2. Arya, V., Garg, N., Khandekar, R., Meyerson, A., Munagala, K., Pandit, V.: Local search heuristics for k-median and facility location problems. SIAM J. Comput. 33(3), 544–562 (2004)
3. Asuncion, A., Newman, D.: UCI machine learning repository (2007), http://www.ics.uci.edu/~mlearn/MLRepository.html
4. Balcan, M.-F., Blum, A., Gupta, A.: Approximate clustering without the approximation. In: SODA '09: 19th Annual ACM -SIAM Symposium on Discrete Algorithms, pp. 1068–1077 (2009)
5. Balcan, M.-F., Blum, A., Vempala, S.: A discriminative framework for clustering via similarity functions. In: STOC 2008: 40th Annual ACM Symposium on Theory of Computing, pp. 671–680 (2008)
6. Balcan, M.-F., Braverman, M.: Finding low error clusterings. In: COLT 2009: 22nd Annual Conference on Learning Theory (2009)
7. Balcan, M.-F., Röglin, H., Teng, S.-H.: Agnostic clustering. In: Gavaldà, R., Lugosi, G., Zeugmann, T., Zilles, S. (eds.) ALT 2009. LNCS, vol. 5809, pp. 384–398. Springer, Heidelberg (2009)
8. Beasley, J.E.: A note on solving large p-median problems. European Journal of Operational Research 21(2), 270–273 (1985)
9. Beasley, J.E.: OR-Library p-median - uncapacitated (1985), http://people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html
10. Bilu, Y., Linial, N.: Are stable instances easy. In: ICS 2010: The First Symposium on Innovations in Computer Science, pp. 332–341 (2010)
11. Gupta, A.: Personal Communication (2009)
12. Jain, K., Mahdian, M., Markakis, E., Saberi, A., Vazirani, V.V.: Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. J. ACM 50(6), 795–824 (2003)
13. Jain, K., Vazirani, V.V.: Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. J. ACM 48(2), 274–296 (2001)
14. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory 28(2), 129–137 (1982)
15. Ostrovsky, R., Rabani, Y., Schulman, L.J., Swamy, C.: The effectiveness of Lloyd-type methods for the k-means problem. In: FOCS '06:47th Annual IEEE Symposium on Foundations of Computer Science, pp. 165–176 (2006)
16. Schalekamp, F., Yu, M., van Zuylen, A.: Clustering with or without the approximation, http://www.itcs.tsinghua.edu.cn/~frans/pub/ClustCOCOON.pdf