Efficient Clustering Schemes Towards Information Collection

Krzysztof Cichoń(✉) and Adrian Kliks

Institute of Radiocommunications, Poznan University of Technology, Poznan, Poland
{krzysztof.cichon,adrian.kliks}@put.poznan.pl

Abstract. One of the challenges in cooperative spectrum sensing is to optimize the energy consumption of the network. Delivery of all measurements from all the nodes to the fusion centre is not the best solution from the perspective of energy-efficiency. Clustering of nodes with similar channel conditions may reduce the amount of transmitted data, and in consequence reduce the amount of consumed energy. In this paper we investigate the performance of selected algorithms known in the domain of artificial intelligence, applied to perform reliable yet energy-aware spectrum sensing.

Keywords: Signal detection · Clustering · Classification

1 Introduction

One of the key assumptions in the cognitive radio networks (CRN) is to sufficiently protect the primary user (PU) transmissions while utilizing the spectrum resources in the flexible way. Operational wireless systems deployed over certain area cannot be disrupted by the secondary user (SU), at most some small and acceptable amount of interference power can be induced [1]. In consequence, the SU has to be aware about the presence of the existing transmissions, and one way to achieve it is to observe - *sense* - the spectrum in order to detect any human-made signals.

Since the origin of the cognitive radio technology (over two decades ago [3]), numerous papers have been published, where researchers investigated various aspects of the spectrum sensing procedure [2]. Interesting survey on the spectrum sensing techniques may be found in [4,5], where the authors reviewed the performance of numerous algorithms designed to spectrum occupancy detection. However, it has been stated that the single-node spectrum sensing (i.e., the procedure where the SU observes spectrum solely without any help or information exchange with other terminals or sensors) cannot guarantee high reliability and accuracy regardless of observation time, as such an approach is vulnerable to e.g. hidden node problem [6]. In this scenario, given SU terminal will be not able to detect the presence of nearby PU due to high shadowing of the signal due to the presence of obstacle (e.g. building). In consequence, single-node spectrum sensing has to be either complemented with the information delivered from the

G. Caso et al. (Eds.): CrownCom 2020, LNICST 374, pp. 45–58, 2021.
https://doi.org/10.1007/978-3-030-73423-7_4

dedicated geolocation databases or collaborative (cooperative) solutions have to be considered. In the latter case even great number of nodes may exchange information to better detect the presence of PU [7]. However, the presence and cooperation between numerous sensors in the system entails the increased energy consumption by the whole system, as well as the increase in latency in decision making. It is due to the need for often periodic delivery of measurement updates from all or most of the cooperating sensors. There is then a trade-off between the number of nodes used to increase the awareness about the surrounding radio environment, and the consumed energy and introduced delay. As sensing is an inherent feature of SU node, the overall energy optimization may be achieved by proper grouping of sensing nodes. Within such a group, the information is processed locally (e.g. the local decision on the spectrum occupancy is made) and later delivered by the group leader to the dedicated sink-node in the network, called often a fusion centre (FC). The FC node collects the messages delivered from all associated clusters, analyse them, perform decisions and distribute them among these clusters.

However, in practice, the locations of the sensing nodes (including wireless user terminals) is not known a-priori, and has to be learnt. It means that the clusters of SU terminals have to be created dynamically, in autonomous and distributed way, as well as the cluster-heads have to be selected in similar way. Clustering of nodes is widely applied and well investigated research problem present in various domains of wireless communications, e.g. [8,9]. In advanced scenarios, where e.g. the reliability of sensed data is not stable, the algorithms known from the domain of artificial intelligence (AI) can be applied. In particular, the clustering algorithms may be considered as efficient tools for node grouping in CRN. Being an extension of the previous paper [10], in this work we compare the proposed soft-K-means algorithms with other AI solutions, properly adjusted to the investigated scenario and applied system model. The rest of the paper is organized as follows. In Sect. 2, the system model and the research problem is presented, and the basics of cooperative spectrum sensing is overviewed. In Sect. 3, the details of considered clustering algorithms are described. Section 4 deals with results analysis, and the whole work is concluded in Sect. 5.

2 Energy Efficiency in Cooperative Spectrum Sensing

In our scheme we consider the set of N wireless nodes capable in performing spectrum sensing in a cooperative way. In order to optimize the energy consumption in the whole network, dynamic node clustering is considered, where cluster representatives (cluster heads, CH) are selected in a dynamic way. Once all the CH are identified, they communicate with the dedicated central node, FC, in order to deliver to it all local information and allow it to make reliable global decision about the spectrum occupancy. As discussed later, the CHs may either send the unprocessed (raw) data gathered from all nodes belonging to the cluster, slightly processed data or even the local decisions. Without loss of generality, we assume that the messages between the nodes are transmitted applying simple binary modulation scheme, i.e. binary phase shift keying (BPSK). The considered scenario is illustrated in Fig. 1.

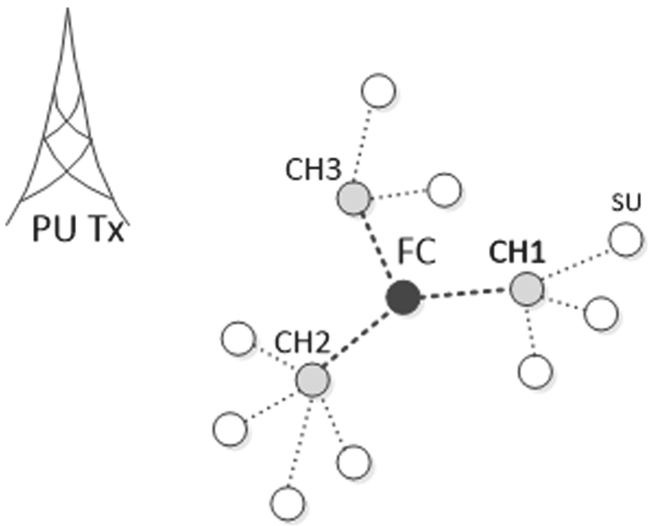

Fig. 1. System model where SUs are sensing PU's signal and are clustered into three clusters (PU - Primary User, SU - Secondary User, CH - Cluster Head, FC - Fusion Centre)

The ultimate goal of any spectrum sensing algorithm is to correctly detect the presence or absence of the signal in the observed band at the specific location. Numerous single-node spectrum sensing solutions have been proposed in the rich literature, starting from the simple yet highly inaccurate in poor channel conditions algorithm known as energy detection, through more advanced eigenvalue detectors or covariance based algorithms, finishing at highly complicated matched-filter schemes [4,5]. However, it has been proved that none of these solutions can guarantee high performance in specific yet common situation, e.g., in case of hidden node problem. In consequence, single node spectrum sensing is said to be not reliable enough to allow for secondary transmission is real-world applications. Contrarily, when the neighbouring nodes could communicate and exchange information, the level of the network awareness on the surrounding radio environment increases.

There are four cases that may happen when the node performs sensing - and all of them constitutes the so-called confusion matrix depicted in Fig. 2. When the node detects the signal if it is indeed presence, one may speak of probability of detection or true-positive scheme. Next, when the channel vacancy is decided when there is no real signal, one may thing on negative-positive scheme. Finally, there are two kinds of wrong decisions - classified as miss-detection (when the real signal is present but is not detected by the SU, false negative case) and classified as probability of false-alarm - the detector decides on signal presence where in reality there is none (false positive case). This decision making process is often referred to hypothesis testing, where absence of PU signal is treated as the null-hypothesis [11,12].

In cooperative case, the nodes transmit their data to the selected cluster head, which is responsible for their processing and final decision making. The nodes may transmit raw data or some initial decisions to the cluster head, which in turns transmit a joint message to the global fusion centre (one dedicated node) to make global decision. In other approach, once the cluster representative node is selected for a certain group of nodes, only this node is responsible for delivering sensing information to the FC. When the fusion centre receives all messages from associated sensing nodes, it may perform final decision by merging delivered information, applying one of the possible rules (OR, AND or K-out-of-N [7]).

		Actual state	
		PU present	PU absent
Detector decision	PU present	Correct positive detection	False alarm
	PU absent	Miss-detection	Correct negative decision

Fig. 2. Confusion matrix of the predicted and actual state in spectrum sensing

Thus, also from this perspective it is reasonable to consider node grouping. Various criteria may be applied in this respect, for example, one may propose to group together the nodes which are geographically close to each other and create clusters based on density of nodes in various areas. Another approach is to group all nodes with similar channel conditions. These issue will be discussed in detail in the following section. Moreover, from the perspective of energy consumption in the CRN, when the total number of cooperating nodes increases, the whole consumed energy increases as well. One may observe that in general, the longer the distance between the nodes, the greater the transmit power necessary to deliver the message to the destination node, and in consequence greater the total energy consumption. In order to evaluate the energy-efficient solutions in CRN there is a need for proper power consumption models, as presented in [10]. The total power consumed by N network members is given by the following formula:

$$P_{\text{total}} = \sum_{i=1}^{N} \left(P_{\text{sens}}^{(i)} + P_{\text{rep}}^{(i)} \right) + P_{\text{shar}}, \tag{1}$$

where $P_{\text{sens}}^{(i)}$ is the power related to spectrum sensing for SU i, which is a constant value for the selected sensing technique and depends on the complexity of the sensing technique. P_{shar} is the power devoted to information sharing within the network, $P_{\text{rep}}^{(i)}$ is the power related to reporting of SU's observation to the cluster head by node i or the power of reporting by the cluster head to Fusion Centre. Reporting power $P_{\text{rep}}^{(i)}$ depends on the multipath loss $\eta^{(i)}$, as well as on the distance between the i-th node and cluster head $d_{\text{CH}}^{(i)}$:

$$P_{\text{rep}}^{(i)} = \frac{\gamma_{target}^{\text{BPSK}} \left(d_{\text{CH}}^{(i)}\right)^{\xi} \sigma_n^2}{\eta^{(i)} G_{\text{t}} G_{\text{r}} \left(\frac{\lambda}{4\pi}\right)^2}, \tag{2}$$

where $\gamma_{target}^{\text{BPSK}}$ is the target signal-to-noise ratio (SNR) for a BPSK-modulated message, ξ is an exponent of the received power decrease dependent on the type of wireless environment, σ_n^2 is the power of noise, G_{t} and G_{r} are the gains of the transmitting and receiving antennas, respectively, and $\frac{\lambda}{4\pi}$ is the wave number [13].

Knowing the energy consumption model, one may formulate the investigation problem, i.e., how to properly cluster the nodes based on the similarity of the observed channel conditions in order to guarantee high energy-efficiency of the entire network. Various algorithms may be applied here, as discussed in the following section.

3 Clustering Algorithms

As specified in the previous sections, we consider the scheme where every SU collects information from the surrounding environment, and delivers it to the cluster head. However, from the network perspective the number of possible clusters as well as the selection of the cluster leader have great impact on the overall energy efficiency. Various clustering criteria may be identified, such as grouping nodes which are close to each other in Euclidean sense. In our scheme, we extend this approach by clustering the nodes which observe similar channel conditions. In other words, when there is a set of nodes even closely deployed, but portion of them is subject to some strong shadowing (with respect to the fusion centre), at least two clusters will be created - one for shadowed and one for non-shadowed group of SUs. In our study we compare some AI based solutions for node clustering (see e.g. [14]). The exemplary clustering results are presented for all algorithms in Fig. 3.

Standard *K*-means. K-means is the well-known clustering scheme in which the target number of clusters K has to be known before the start of the algorithm [15]. In short, the algorithm operates in iterative mode - in the first *assignment* step, each SU is assigned to the closest centroid (cluster centre). In second step the positions of centroids are updated on the base of assigned cluster members. Thus, the second phase is called an *update step*.

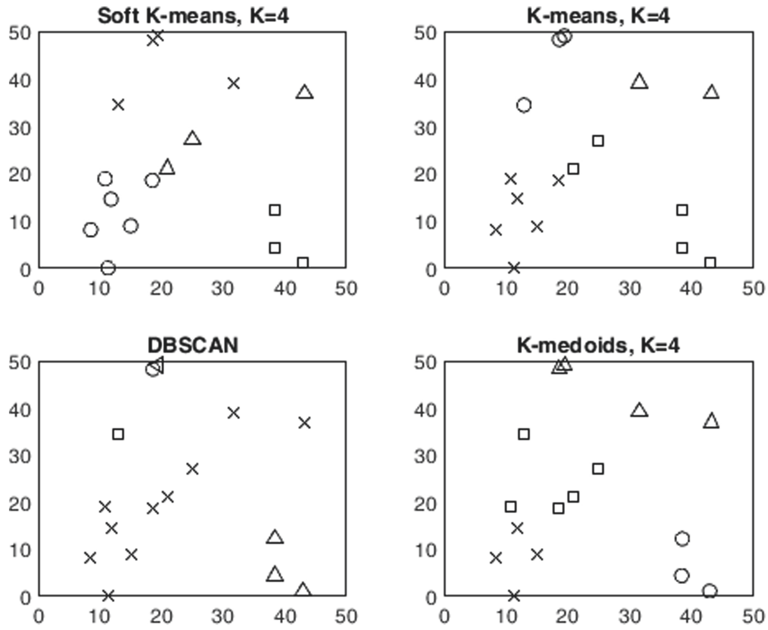

Fig. 3. Exemplary clustering results for $N = 16$ SUs for clustering methods considered in the article. The same node distribution is applied to find the members of clusters (distinguished with various shapes: squares, triangles, circles, etc.)

Soft $\boldsymbol{K}$**-means.** In the soft K-means the way how the nodes are assigned to clusters is *soft* what means that the node can be *partially* member of more than one cluster. The *degree of assignment* is then given by the formula:

$$r_k^{(n)} = \frac{\exp\left(-\theta d_{m^{(k)}}^{(i)}\right)}{\sum_{k'} \exp\left(-\theta d_{m^{(k')}}^{(i)}\right)}, \tag{3}$$

where $d_{m^{(k)}}^{(i)}$ is the distance between SU i and the centroid $m^{(k)}$ of cluster k. The sum of all assignments for the nth SU to all centroids considered in the network is one. θ is the stiffness which is the key parameter for the algorithm and is an inverse-length-squared.

$\boldsymbol{K}$**-medoids.** In K-medoids the clustering is considered to be similar to K-means article. In the latter, the new centroid is found as the mean of locations of cluster members. The algorithm may be vulnerable to outlier nodes which are distant from main cluster members and may disturb the correct location of cluster head. However, different than in K-means, the location of the medoid is always the position of the real node, since each SU is tested as the potential medoid. Moreover, in K-medoids the central point of cluster is selected on the base of the closest distance to remaining potential cluster members. The procedure is iteratively conducted for any network member.

Density-Based Approach. Density-based spatial clustering, known as DBSCAN, is the algorithm where the number of clusters is not needed to be specified at the beginning of the algorithm. Instead, two other parameters are required: ϵ, being the radius of the neighbourhood under consideration, and minimum cluster size. First, the neighbours are found within the ϵ neighbourhood. If the number of neighbours is lower than minimum cluster size, then these nodes become so-called *noise nodes*, otherwise they form the cluster and become *clustered nodes*. The cluster could be extended after the time the procedure is repeated for border members of newly formed cluster. If no new cluster members could be added, the procedure is repeated for other unlabelled nodes with the possible formation of different cluster. The key parameter in the algorithm is the selection of the ϵ value which is the radius around the point which is used for adding new cluster members. Moreover, the minimum cluster size is the second important parameter which should be no lower than 3.

4 Simulation Results

Extensive simulations have been carried out toward reliable performance comparison of the selected well-known clustering algorithms, i.e., K-means, K-medoids and DBSCAN with the proposed soft-K-means solution [10]. In our simulation we consider the set of N from 8 to 20 nodes deployed uniformly over the considered geographical area of size 50 m. The nodes can communicate wirelessly by sending BPSK modulated signals. The required SNR that has to be guaranteed at the receiving nodes was set to $\gamma_{target}^{\mathrm{BPSK}} = 11.3$ which is found for target BER equal to 10^{-6}. The value of the channel decaying factor ξ was equal to 3.5. The transmit and receive antenna gains have been setup to $G_{\mathrm{t}} = 1$ and $G_{\mathrm{r}} = 1$, respectively, and the considered wavelength was found for carrier frequency of 2.4 GHz $\lambda = \frac{c}{f_c} = 0.125$ m.

In the presented results two reference cases without clustering are considered. In the first one, called as ‘SNR-based selection’, a majority of nodes report their observation to FC, the nodes with the highest SNR are selected. In ‘energy-based selection’ the same number of nodes is selected, based on the lowest energy usage in reporting link. In the following section the ‘reporting node’ is the node which reports sensing observations directly to FC, regardless of the clustering procedure is applied (DBSCAN, K-means, K-medoids) or not (SNR- and energy based selections).

First, DBSCAN algorithm is investigated for finding such set of parameters which guarantee efficient operation. In Fig. 4, the consumed energy of the network is presented as the function of the number of nodes. One can observe that the minimum energy (given in Joules) is observed for various DBSCAN algorithms. First, the consumed energy increases with the number of nodes for any variant of DBSCAN algorithm. And for 8–10 nodes the lowest energy usage is for $\epsilon = 16$, but this case for the network of 20 nodes becomes inefficient. Then the trend is the following: the greater the number of nodes in the network, the lower the optimum value of ϵ: for example for 14–18 nodes the lowest energy usage is observed for $\epsilon = 12$ and for 20 nodes for $\epsilon = 10$. The reference cases, i.e. energy selection and SNR-based selection (for significant number of nodes), offer much greater energy consumption than DBSCAN grouping.

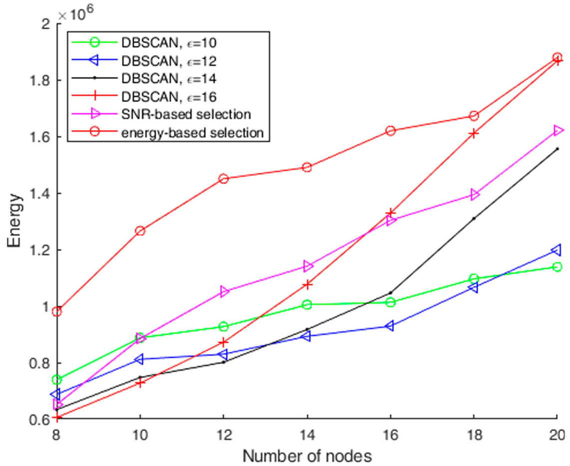

Fig. 4. Energy usage by DBSCAN algorithm with various values ϵ

This is mainly due to the greater energy consumption for reporting links from SUs to Fusion Centre than to cluster heads. In Fig. 5 one can see the number of reporting nodes which for reference cases is significantly greater than for DBSCAN algorithms. This is because in reference cases clusters are not formed, the SUs are independent and report observation directly to Fusion Centre. However, in DBSCAN parameter ϵ discriminates the number of reporting nodes (which is the number of clusters and *noise* nodes. With the increasing number of nodes the number of reporting links is slightly going down. The reason for this is that the greater number of nodes, the lower number of possible *noise* nodes and it is more probable to form the cluster with minimum number of nodes required in the algorithm.

In Fig. 6 the energy usage for K-medoids is shown for various number of clusters K. The lowest energy consumption is observed for two cases: for 4 and 5 clusters. The energy usage raises with number of nodes but is still significantly lower than for reference cases (with the exception of SNR-based scheme for low number of nodes).

Then, in Fig. 7 one can see the energy usage for standard K-means scheme. The energy usage is similar within a couple of cases where the number of clusters is from 3 to 5. The case with $K = 2$ clusters offer higher energy usage what makes it unuseful.

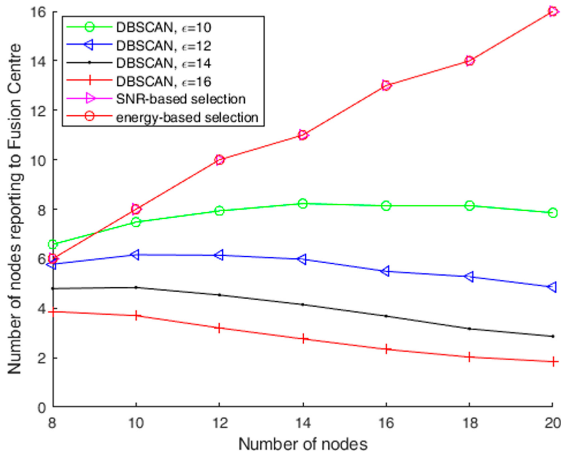

Fig. 5. Number of reporting nodes for DBSCAN algorithm with various values ϵ

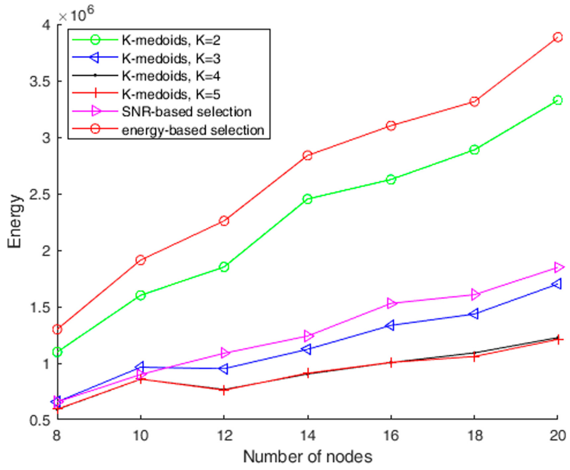

Fig. 6. Energy consumption for the whole network within K-medoids algorithm

The comparison of energy consumption for soft K-means is highlighted in Fig. 7. The crucial parameter of *stiffness* is set for all cases as 0.15 according to the highlights in [10]. However, energy usage for three cases is similar: $K = \{3, 4, 5\}$. Slightly lower energy usage is observed for $K = 4$ clusters. What is important in soft K-means the energy spent in the sensing process is lower than in standard K-means (Fig. 8).

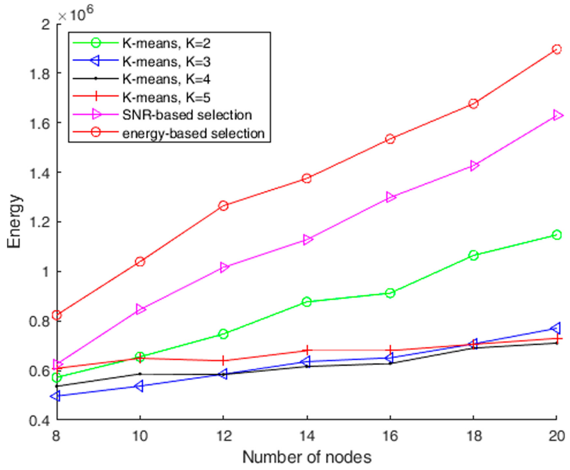

Fig. 7. Energy consumption for the whole network within standard K-means

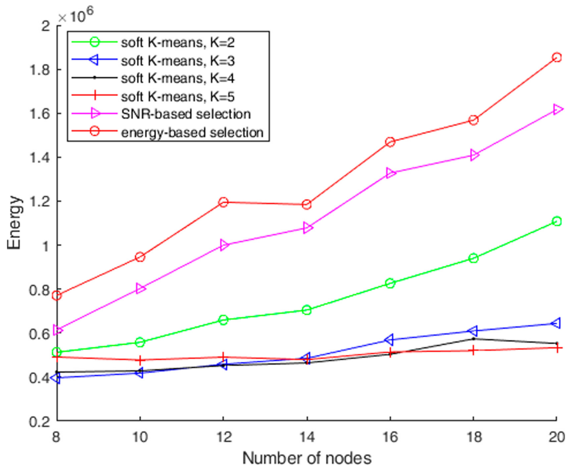

Fig. 8. Energy consumption for the whole network within soft K-means

Finally, the results for all considered algorithms were compared in Fig. 9. The input parameters were selected in the way that 4 clusters is found. In DBSCAN, where the number of clusters cannot be defined a priori since it is the result of ϵ and minimum cluster size parameters, it is around 4 for $\epsilon = 14$. Within such set of algorithms the lowest energy usage is observed for soft K-means. The energy usage is of 0.5 J and increases slightly when more nodes are considered. The second-best is standard K-means offering also good stability of energy consumption versus number of nodes and offering about 35% higher energy usage than soft K-means. K-medoids and DBSCAN offer similar performance to standard K-means when number of nodes is of 8 to 12 nodes. However, for greater network size the consumed energy is far bigger than for K-means schemes. Moreover, the reference cases, based on simple centralised topologies involve significantly higher energy usage than clustered schemes.

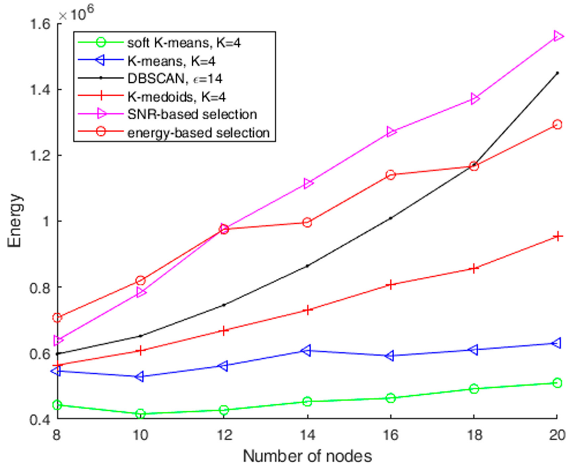

Fig. 9. Energy consumption comparison for all considered grouping algorithms

The reason for this is much greater number of reporting nodes when compared to other grouping schemes, what is included in Fig. 10. Indeed, K-medoids and both versions of K-means have constant number of 4 reporting nodes, in DBSCAN it decreases from 5 to 3, however, in SNR and energy selection number of reporting nodes is the total number of network members and increases from 6 to 16 (around 75% of network members). Therefore, the soft K-means is the most promising solution since due to its soft *stiffness* metric the required number of K clusters may be lowered by merging some clusters according to the procedure. This can be observed in Fig. 10 where the number of reporting nodes for soft K-means is lower than 4. Moreover, it offers the same detection quality (Fig. 11) with lower energy consumption when compared to other clustering solutions. The global detection quality for all clustering schemes remains the same for a given number of nodes and grows when network size increases.

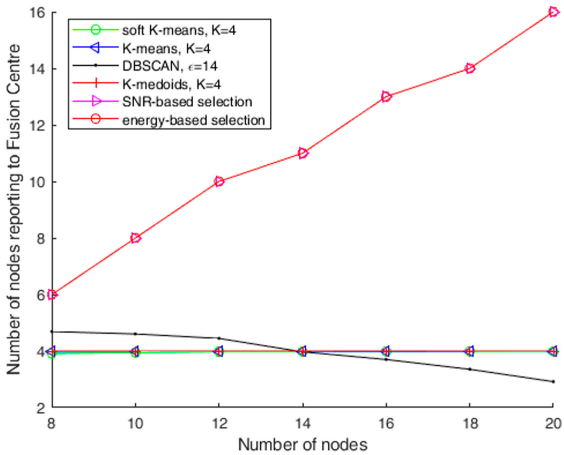

Fig. 10. Number of clusters formed for various clustering schemes

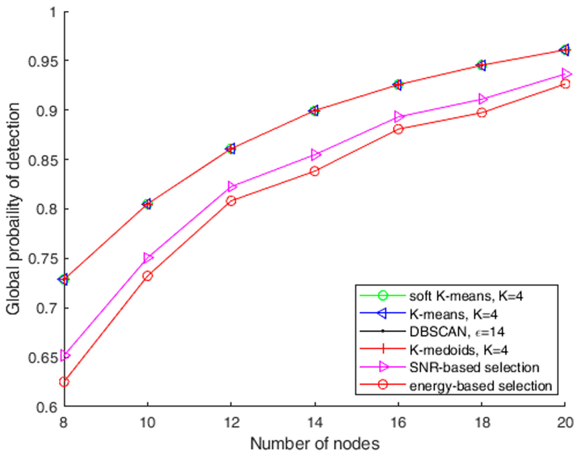

Fig. 11. Global probability of detection for various clustering schemes

5 Conclusions

In this work the analysis of energy efficiency for selected clustering procedures is provided. First, density-based approach with noise nodes is analysed and optimum values for input parameters are found. Then, the energy usage for K-medoids, K-means and proposed soft K-means is validated. The conducted simulations have shown the lowest energy usage is observed for soft K-means scheme. What is more, the detection quality is unchanged. Then, for K-means, the energy usage is higher, but low complexity is a great advantage. Moreover soft and standard K-means both offer the energy usage which is not growing rapidly with number of nodes in the network. In future works it is planned to compare also distributed clustering schemes with the presented hierarchical ones. Moreover, mobility of SUs and network is the important factor for detection quality and energy usage and is planned to be included in the model.

Acknowledgement. The presented work has been funded by the National Science Centre in Poland within the CERTAIN project (no. 2017/27/L/ST7/03166) of the DAINA programme.

References

1. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005). https://doi.org/10.1109/JSAC.2004.839380
2. Haykin, S., Thomson, D.J., Reed, J.H.: Spectrum sensing for cognitive radio. Proc. IEEE **97**(5), 849–877 (2009). https://doi.org/10.1109/JPROC.2009.2015711
3. Mitola, J., Maguire, G.Q.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999). https://doi.org/10.1109/98.788210
4. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009). https://doi.org/10.1109/SURV.2009.090109
5. Ali, A., Hamouda, W.: Advances on spectrum sensing for cognitive radio networks: theory and applications. IEEE Commun. Surv. Tutor. **19**(2), 1277–1304 (2017). https://doi.org/10.1109/COMST.2016.2631080
6. Nekovee, M.: A survey of cognitive radio access to TV white spaces. In: 2009 International Conference on Ultra Modern Telecommunications and Workshops, St. Petersburg, pp. 1–8 (2009). https://doi.org/10.1109/ICUMT.2009.5345318
7. Cichoń, K., Kliks, A., Bogucka, H.: Energy-efficient cooperative spectrum sensing: a survey. IEEE Commun. Surv. Tutor. **18**(3), 1861–1886 (2016). https://doi.org/10.1109/COMST.2016.2553178
8. Din, S., Ahmad, A., Paul, A., Ullah Rathore, M.M., Jeon, G.: A cluster-based data fusion technique to analyze big data in wireless multi-sensor system. IEEE Access **5**, 5069–5083 (2017). https://doi.org/10.1109/ACCESS.2017.2679207
9. Zhang, D., Ge, H., Zhang, T., Cui, Y., Liu, X., Mao, G.: New multi-hop clustering algorithm for vehicular ad hoc networks. IEEE Trans. Intell. Transp. Syst. **20**(4), 1517–1530 (2019). https://doi.org/10.1109/TITS.2018.2853165
10. Cichoń, K.: Soft network organisation towards future distributed ML-Based sensing systems. In: 28th International Conference on Software, Telecommunications and Computer Networks, Hvar (2020). https://doi.org/10.23919/SoftCOM50211.2020.9238189
11. Jayaweera, S.K.: Introduction to detection theory. In: Signal Processing for Cognitive Radios, pp. 65–131. Wiley (2015). https://doi.org/10.1002/9781118824818.ch4
12. Hippenstiel, R.D.: Detection Theory: Applications and Digital Signal Processing, 1st edn. CRC Press, Boca Raton (2001). SBN-10 0849304342. ISBN-13 978-0849304347

13. Cichoń, K., Bogucka, H.: An energy-efficient cooperative spectrum sensing. In: IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp. 24–28 (2015). https://doi.org/10.1109/BlackSeaCom.2015.7185079
14. Xu, R., Wunsch, D.: Clustering. Wiley-IEEE Press, Hoboken (2008). ISBN-10 0470276800. ISBN-13 978-0470276808
15. MacKay, D.J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge (2002)