Supplementary Material to Experimental Evaluation of Szemerédi's Regularity Lemma in Graph-based Clustering - Part I

Jian Hou[a], Juntao Ge[a], Huaqiang Yuan[a], Marcello Pelillo[b,c]

[a]*School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China*
[b]*DAIS, Ca' Foscari University, Venice 30172, Italy*
[c]*European Centre for Living Technology, Ca' Foscari University, Venice 30123, Italy*

Abstract

We introduce a set of experiments which are skipped in the paper for space reason.

1. Experiments

In Section 4 of the paper, some experimental results with the ACC (accuracy), ARI (adjusted rand index) and RI (rand index) criteria are not reported due to limited space. In order to show that the conclusions from the NMI (normalized mutual information) are also applicable to other evaluation criteria, we report the results with ACC and ARI criteria here. The results with RI (rand index) are reported in the second part of the supplementary material.

1.1. Influence of parameters

Using the same experimental setup as in the paper and with the ACC and ARI criteria, the clustering results of the enhanced SPC, APC, DSet and SPRG algorithms with different parameters are reported in Figure 1 to Figure 8. From these tables we observe that, with the ACC and ARI criteria, we arrive at similar observations as with the NMI criteria in Section 4.1 of the paper.

In order to check if the recommended narrow ranges of parameters are effective, we compare the results from wide ranges of parameters with those from narrow ranges of parameters. The comparisons with ACC and ARI criteria are

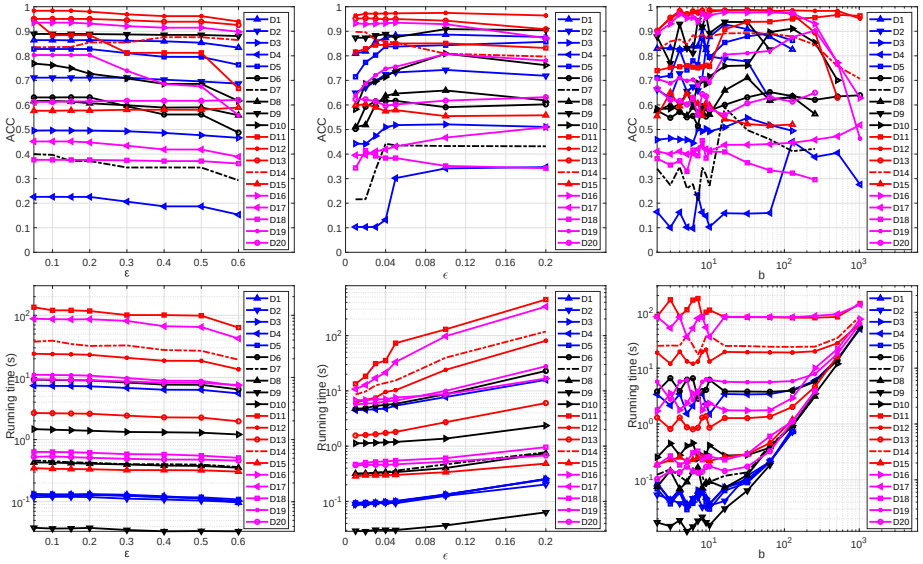

Figure 1: Clustering results (ACC) of the enhanced SPC algorithm, with different parameters.

reported in Figure 9 and Figure 10. It can be observed that with the ACC and ARI criteria, the results from the recommended narrow ranges of parameters are similar to those from wide ranges of parameters.

1.2. Comparison with the original algorithms

In comparison with the original algorithms, we report the clustering results of the enhanced SPC, APC, DSet and SPRG algorithms with the ACC and ARI criteria in Figure 11 to Figure 14. It can be observed that with all the criteria, the enhanced algorithms outperform the original ones in both clustering results and computation efficiency, consistent with the observation in Section 4.2 of the paper.

In order to check if it is necessary to use regularity partitioning in building the reduced graph, we use k-means to partition the graph and compare with regularity partitioning method. The comparisons with ACC and ARI criteria are shown in Figure 15 and Figure 16. It can be observed that with the ACC and ARI criteria, the regularity partitioning outperforms k-means-based partitioning on most of the datasets.

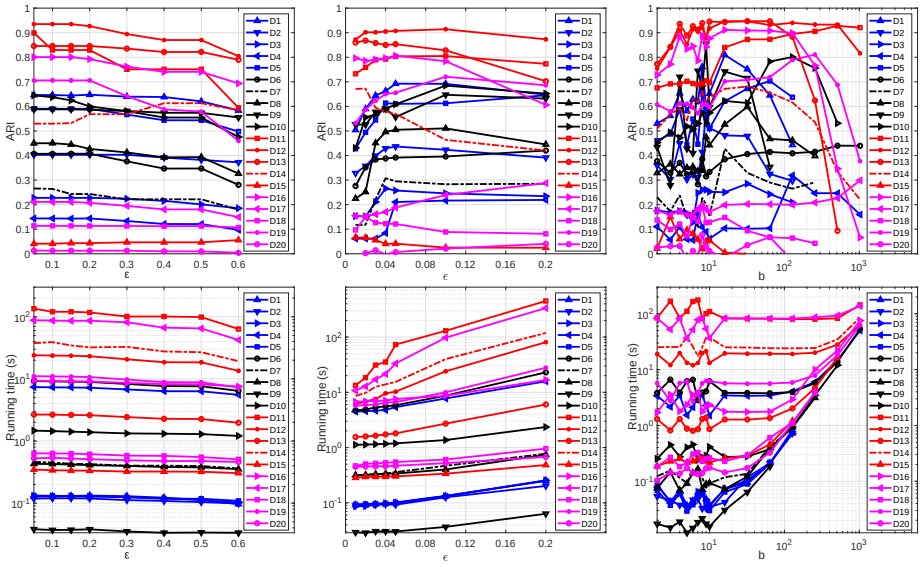

Figure 2: Clustering results (ARI) of the enhanced SPC algorithm, with different parameters.

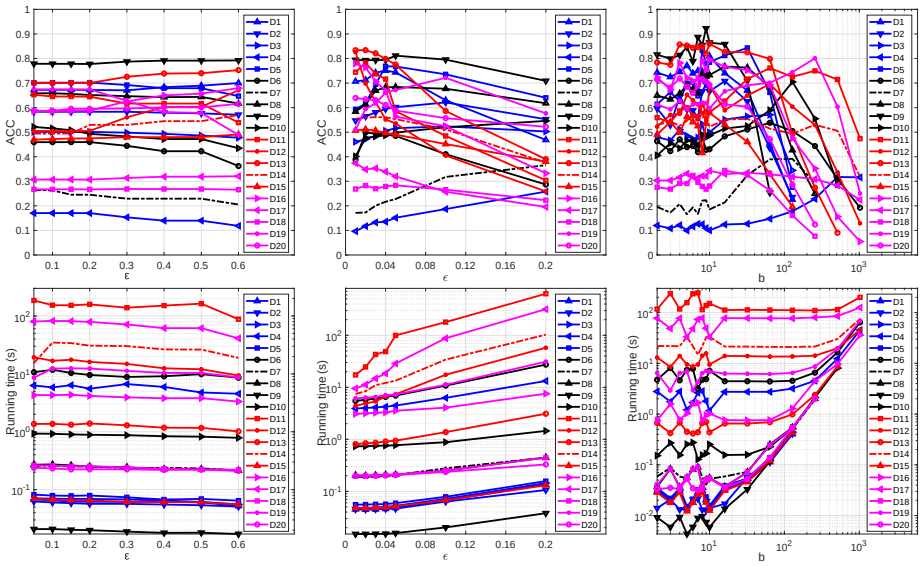

Figure 3: Clustering results (ACC) of the enhanced APC algorithm, with different parameters.

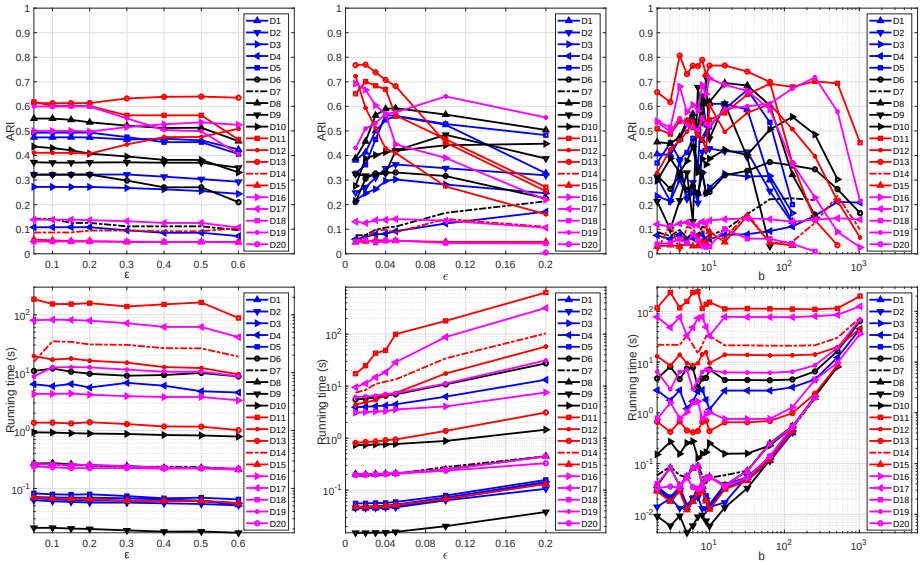

Figure 4: Clustering results (ARI) of the enhanced APC algorithm, with different parameters.

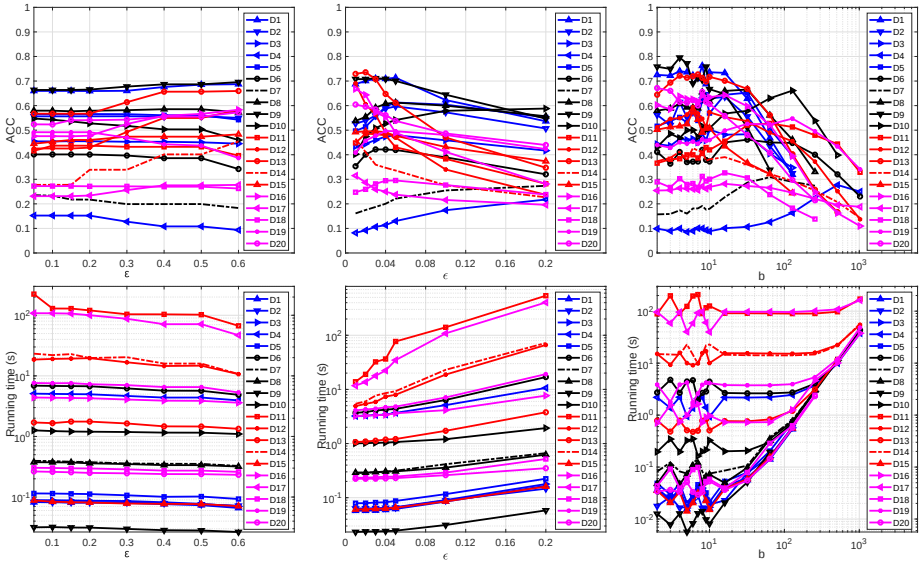

Figure 5: Clustering results (ACC) of the enhanced DSet algorithm, with different parameters.

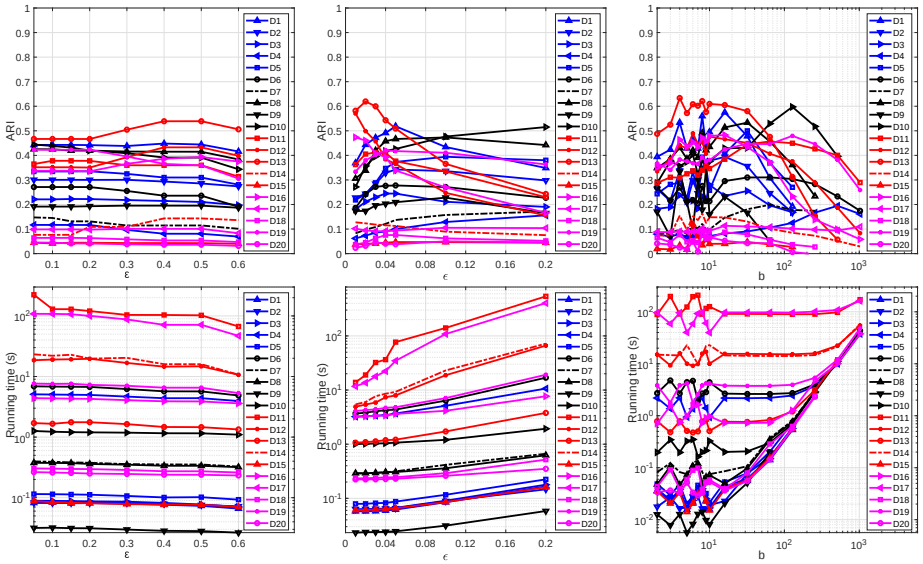

Figure 6: Clustering results (ARI) of the enhanced DSet algorithm, with different parameters.

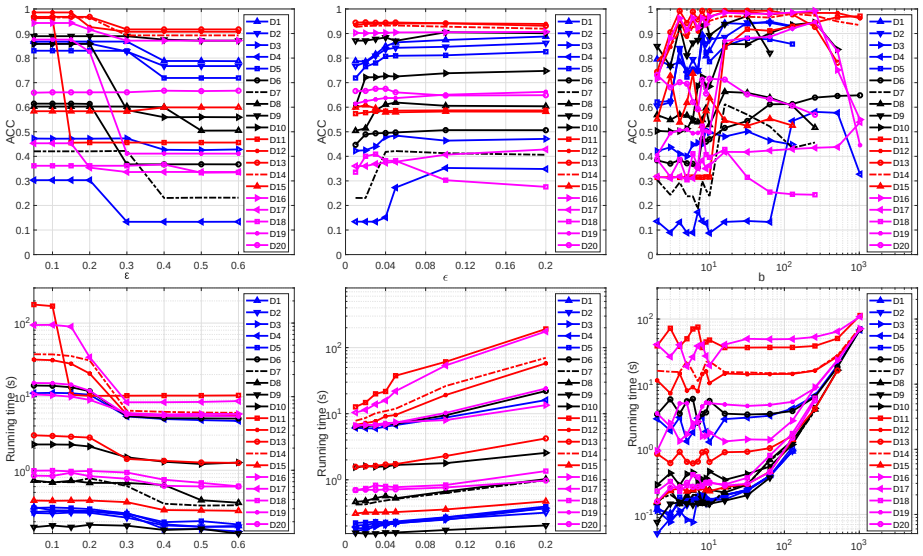

Figure 7: Clustering results (ACC) of the enhanced SPRG algorithm, with different parameters.

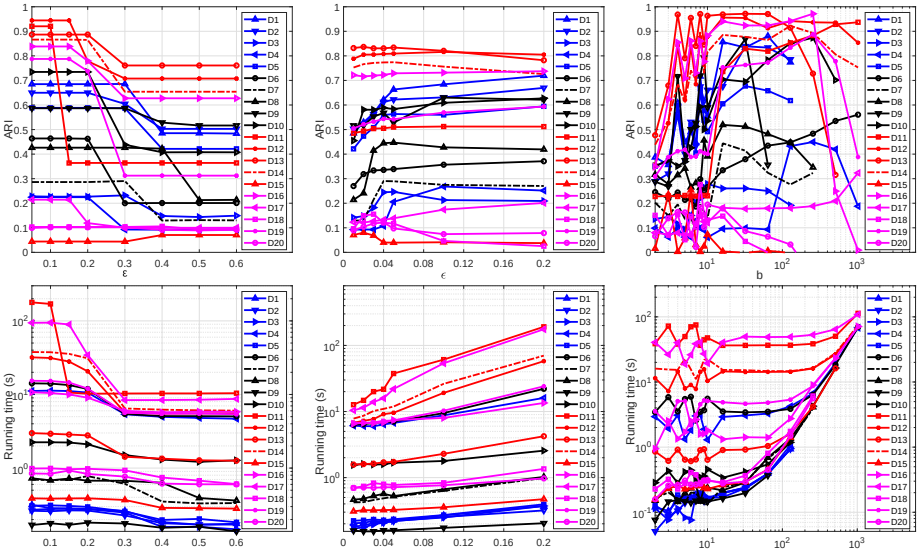

Figure 8: Clustering results (ARI) of the enhanced SPRG algorithm, with different parameters.

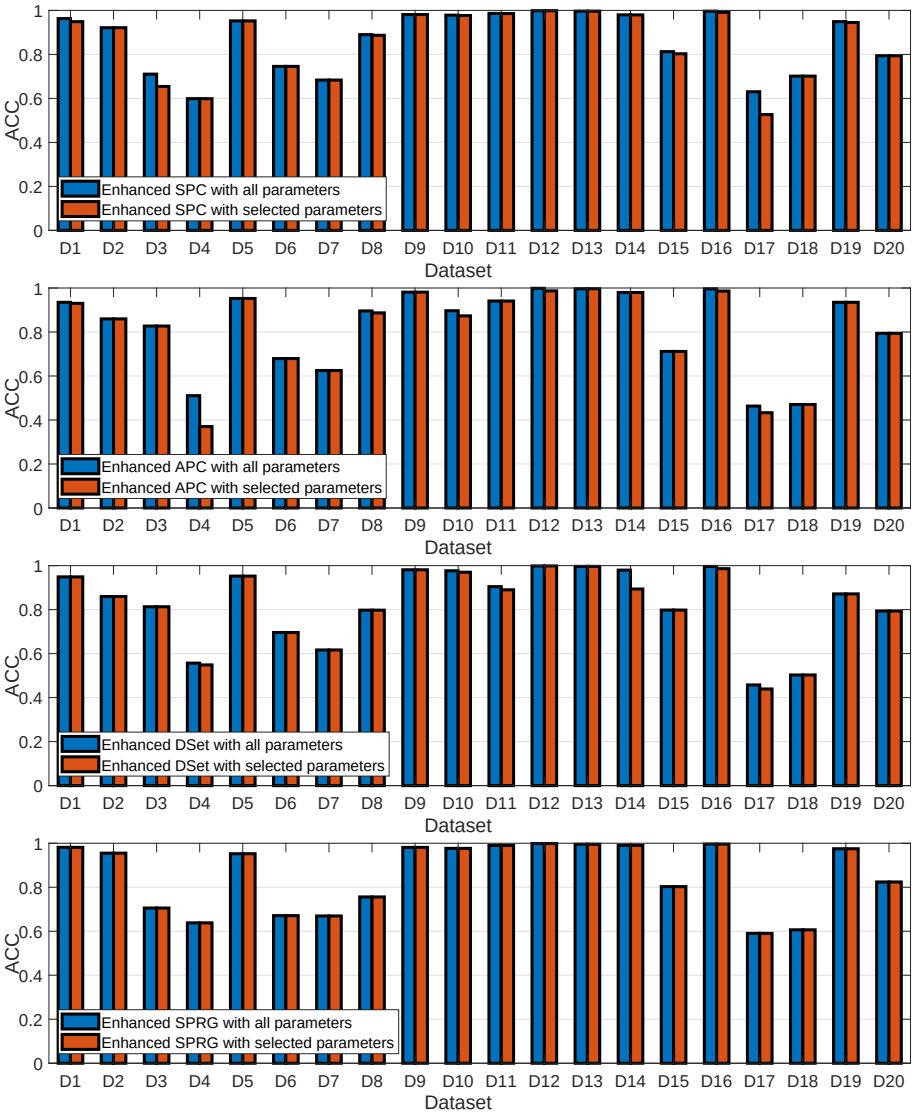

Figure 9: Comparison between the results (ACC) with all parameters and with selected parameters.

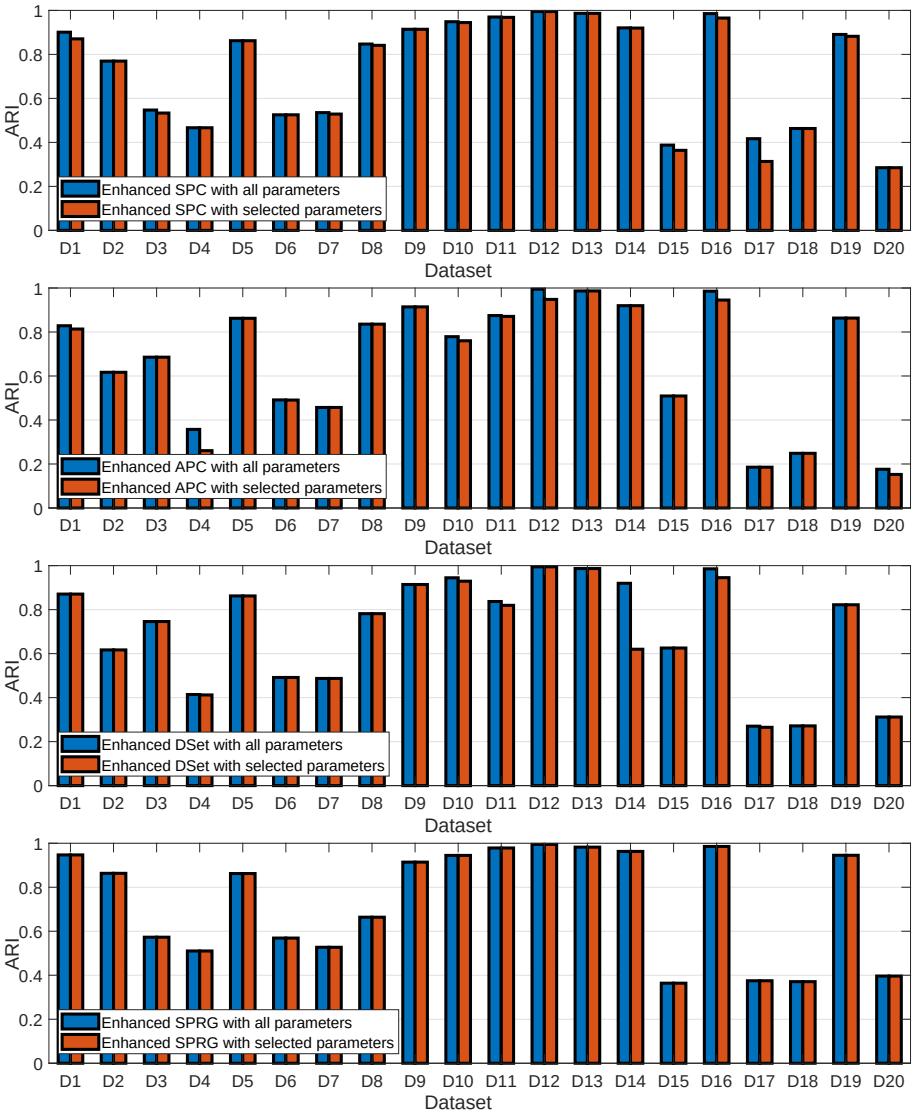

Figure 10: Comparison between the results (ARI) with all parameters and with selected parameters.

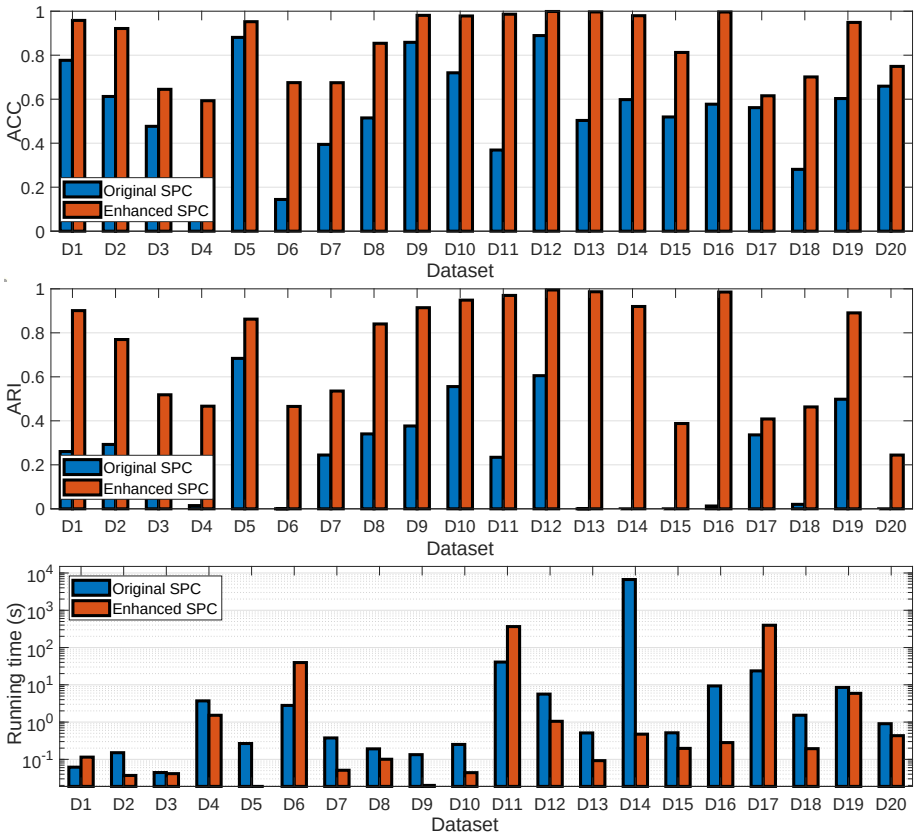

Figure 11: Comparison between the enhanced and original SPC algorithm. The enhanced version generates better results with less running time on the majority of datasets.

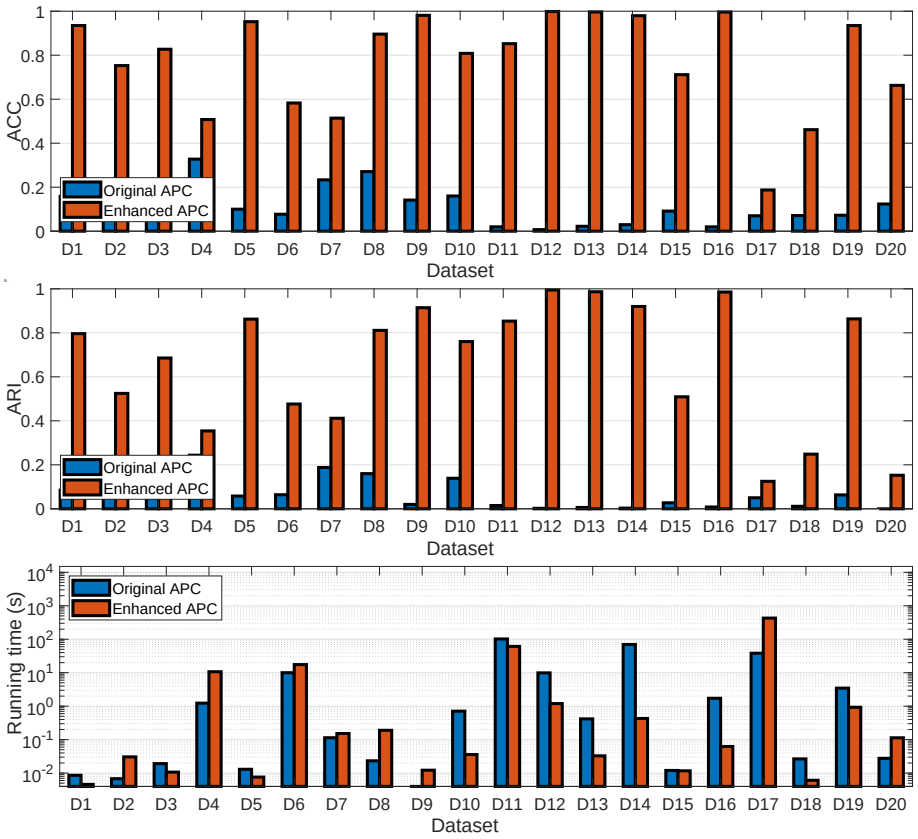

Figure 12: Comparison between the enhanced and original APC algorithme. The enhanced version generates better results with less running time on the majority of datasets.

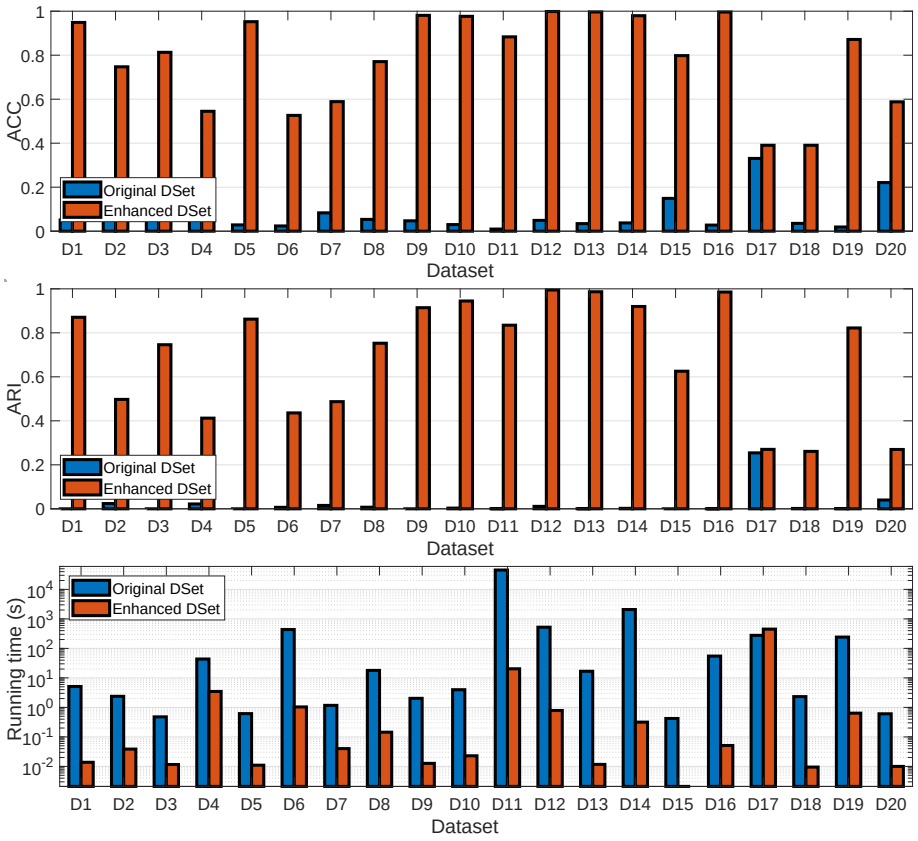

Figure 13: Comparison between the enhanced and original DSet algorithm. The enhanced version generates better results with less running time on the majority of datasets.

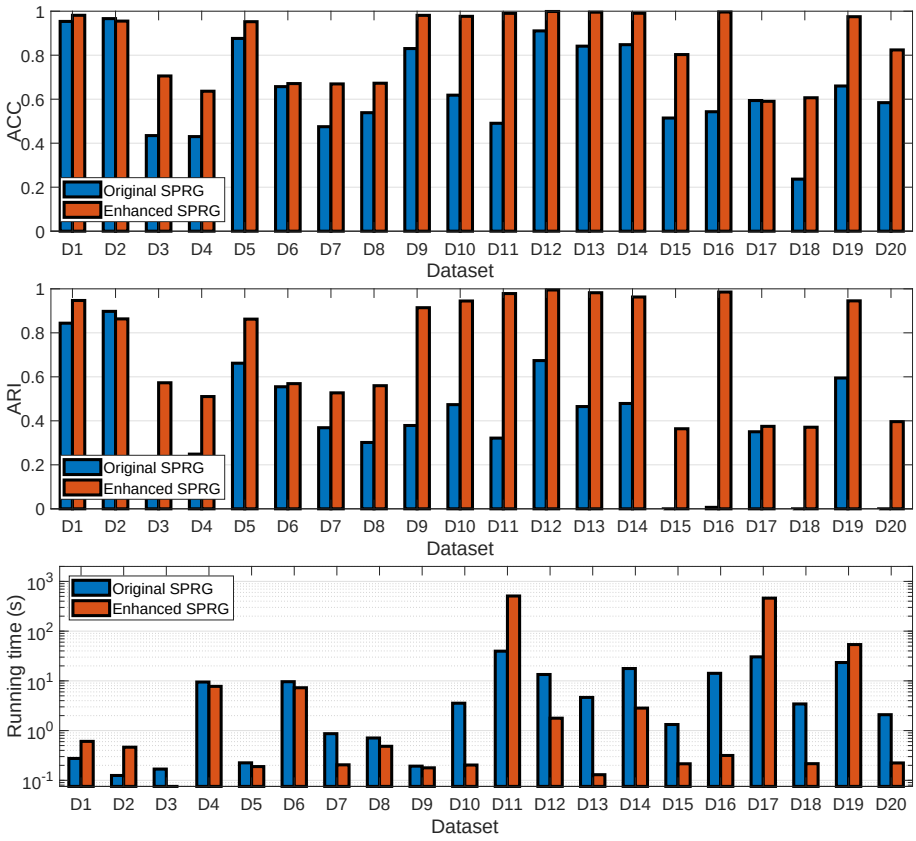

Figure 14: Comparison between the enhanced and original SPRG algorithm. The enhanced version generates better results with less running time on the majority of datasets.

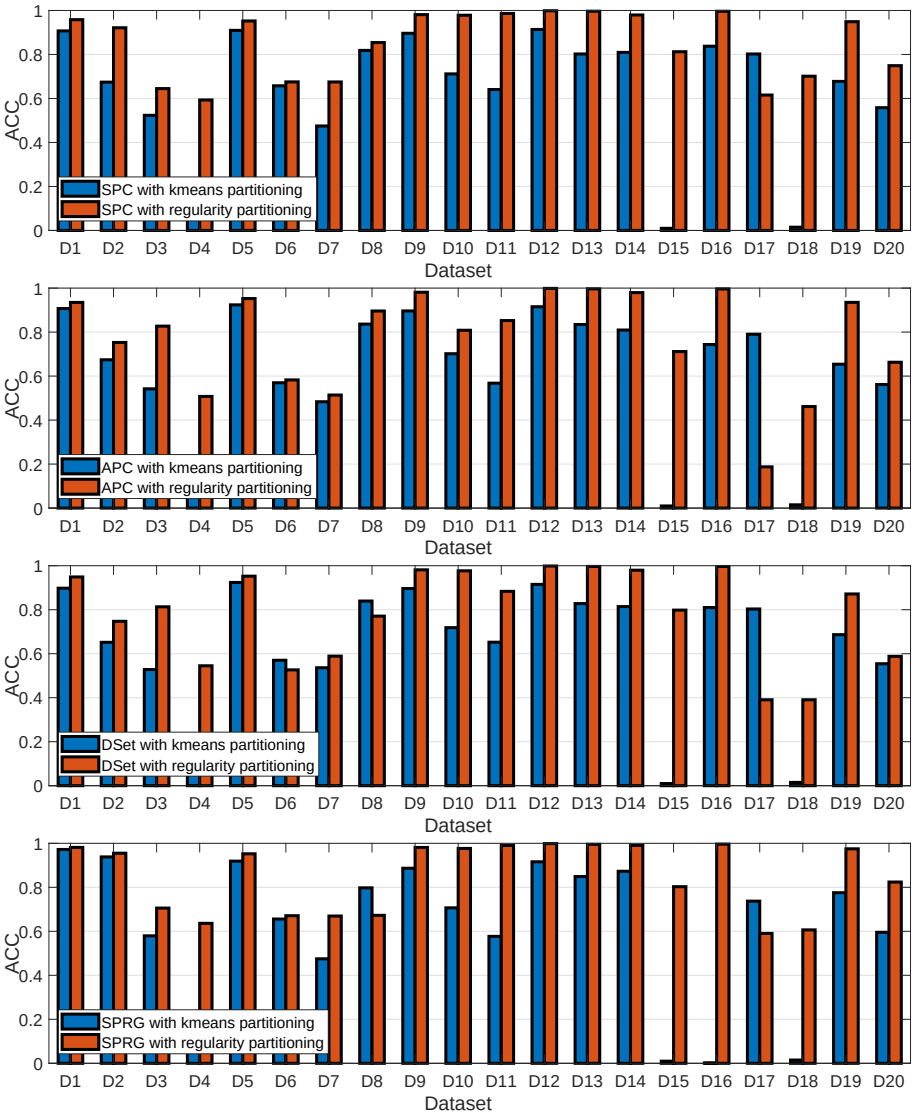

Figure 15: Comparison of ACC between the enhanced algorithms with kmeans-based partitioning and with regularity partitioning. In most cases the regularity partitioning is shown to perform better than kmeans-based partitioning.

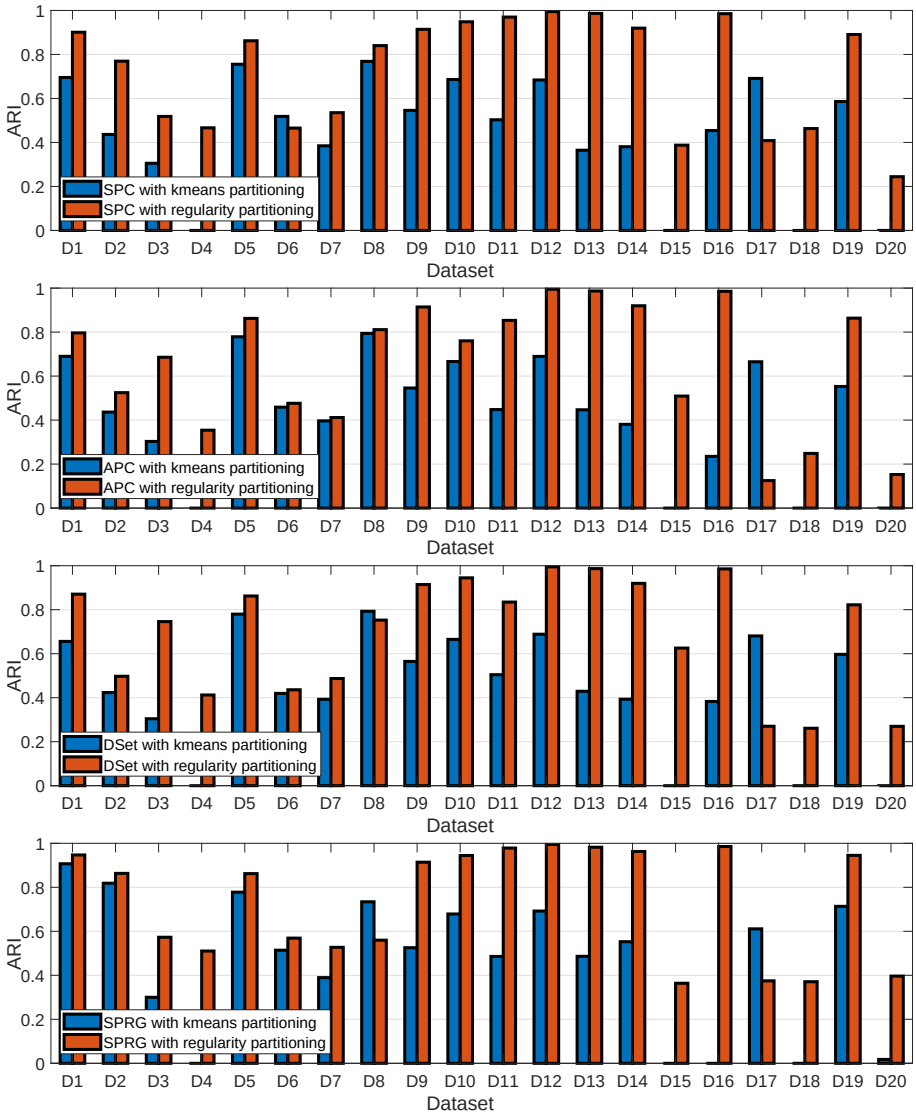

Figure 16: Comparison of ARI between the enhanced algorithms with kmeans-based partitioning and with regularity partitioning. In most cases the regularity partitioning is shown to perform better than kmeans-based partitioning.

Supplementary Material to Experimental Evaluation of Szemerédi's Regularity Lemma in Graph-based Clustering - Part II

Jian Hou[a], Juntao Ge[a], Huaqiang Yuan[a], Marcello Pelillo[b,c]

[a]*School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China*
[b]*DAIS, Ca' Foscari University, Venice 30172, Italy*
[c]*European Centre for Living Technology, Ca' Foscari University, Venice 30123, Italy*

Abstract

We introduce a set of experiments which are skipped in the paper for space reason.

1. Experiments

In Section 4 of the paper, some experimental results with the ACC (accuracy) and ARI (adjusted rand index) criteria are not reported due to limited space. In order to show that the conclusions from the NMI (normalized mutual information) are also applicable to other evaluation criteria, we report the results with the RI criterion here. The results with ACC and ARI are reported in the first part of the supplementary material.

1.1. Influence of parameters

Using the same experimental setup as in the paper and with the RI criterion, the clustering results of the enhanced SPC, APC, DSet and SPRG algorithms with different parameters are reported in Figure 1 to Figure 4. From these tables we observe that, with the RI criteria, we arrive at similar observations as with the NMI criteria in Section 4.1 of the paper.

In order to check if the recommended narrow ranges of parameters are effective, we compare the results from wide ranges of parameters with those from

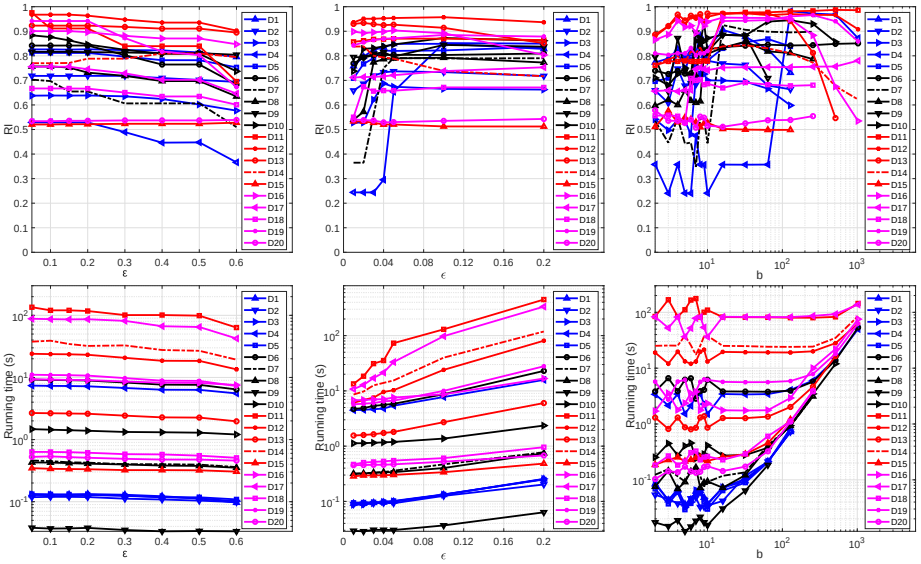

Figure 1: Clustering results (RI) of the enhanced SPC algorithm, with different parameters.

narrow ranges of parameters. The comparisons with the RI criteria are reported in Figure 5. It can be observed that with the RI criterion, the results from the recommended narrow ranges of parameters are similar to those from wide ranges of parameters.

1.2. Comparison with the original algorithms

In comparison with the original algorithms, we report the clustering results of the enhanced SPC, APC, DSet and SPRG algorithms with the RI criterion in Figure 6 to Figure 9. It can be observed that with all the criteria, the enhanced algorithms outperform the original ones in both clustering results and computation efficiency, consistent with the observation in Section 4.2 of the paper.

In order to check if it is necessary to use regularity partitioning in building the reduced graph, we use k-means to partition the graph and compare with regularity partitioning method. The comparisons with the RI criterion are shown in Figure 10. It can be observed that with the RI criterion, the regularity partitioning outperforms k-means-based partitioning on most of the datasets.

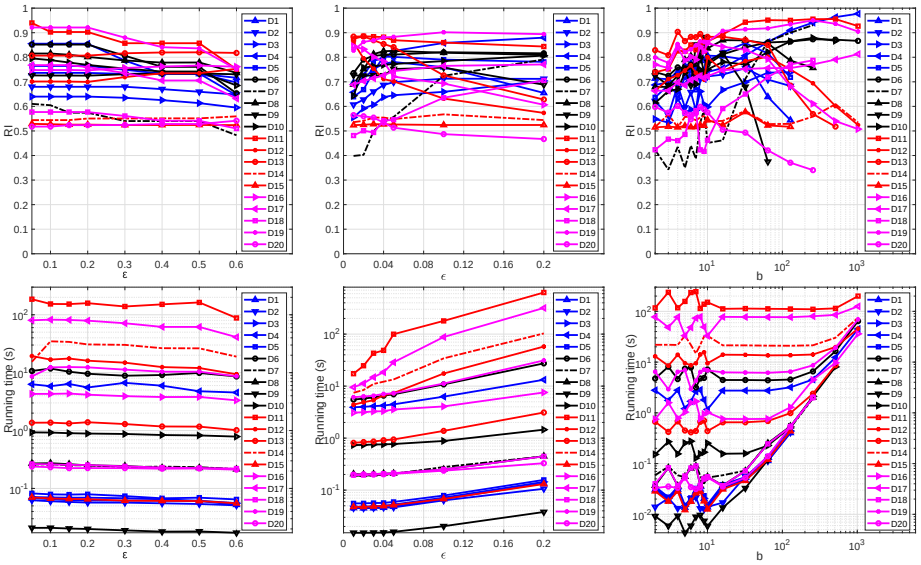

Figure 2: Clustering results (RI) of the enhanced APC algorithm, with different parameters.

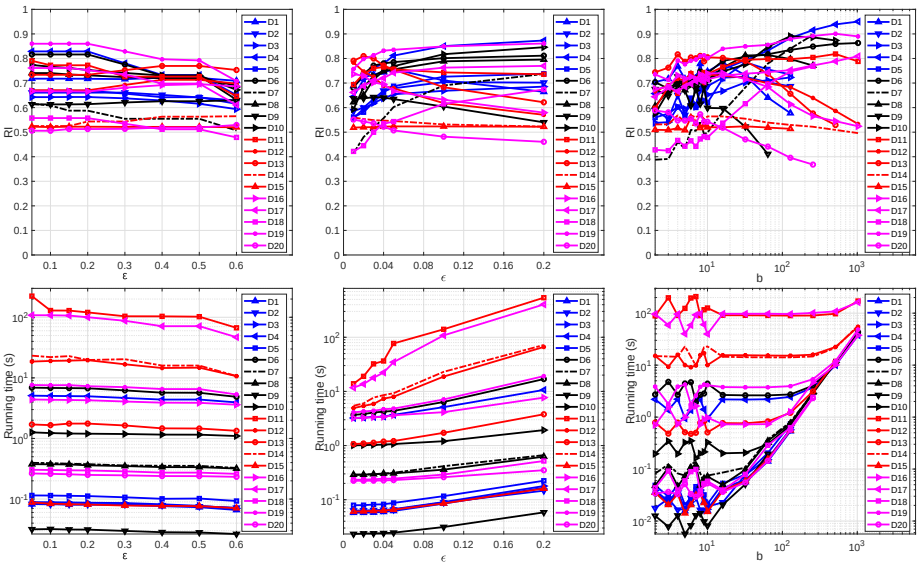

Figure 3: Clustering results (RI) of the enhanced DSet algorithm, with different parameters.

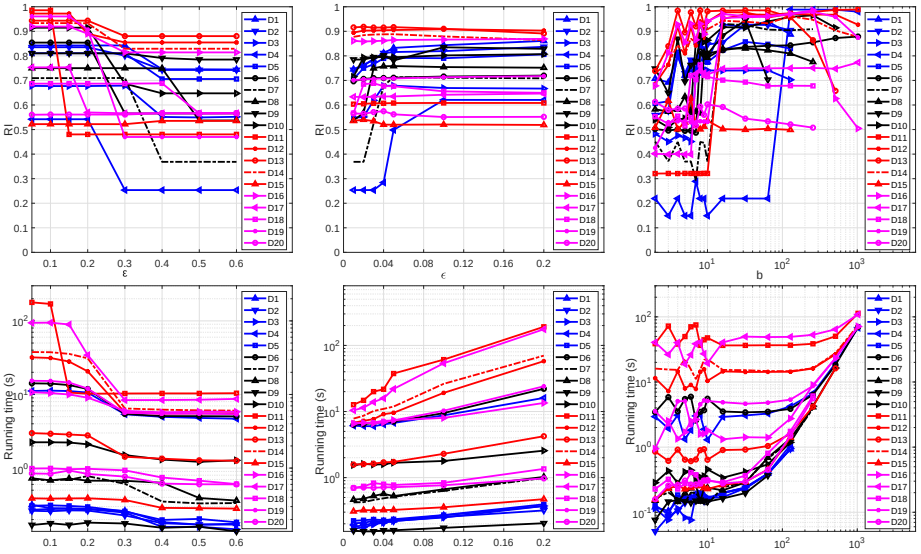

Figure 4: Clustering results (RI) of the enhanced SPRG algorithm, with different parameters.

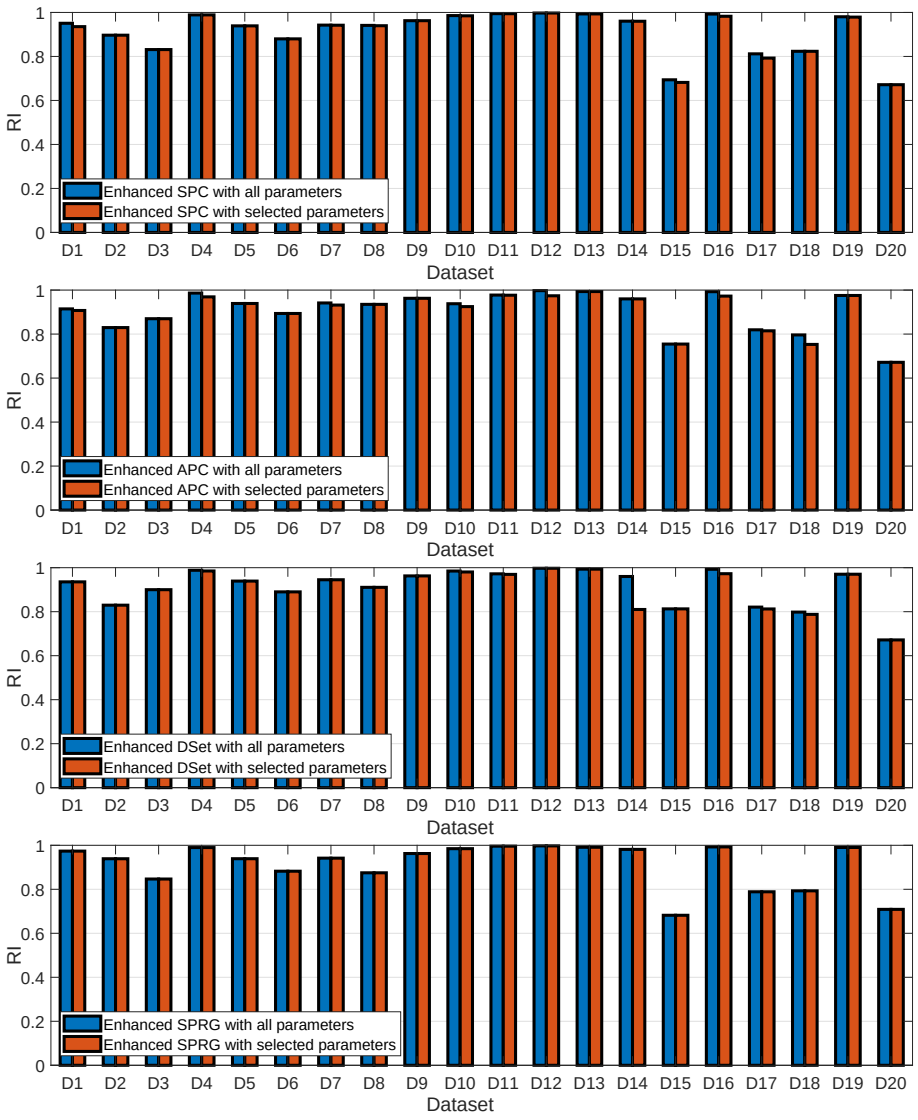

Figure 5: Comparison between the results (RI) with all parameters and with selected parameters.

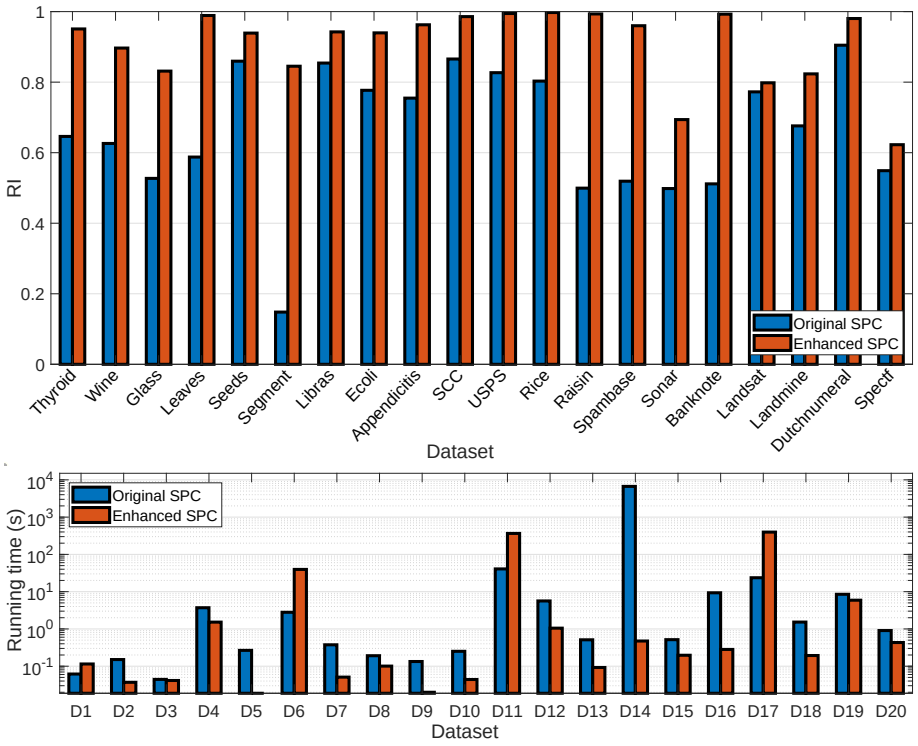

Figure 6: Comparison between the enhanced and original SPC algorithm. The enhanced version generates better results with less running time on the majority of datasets.

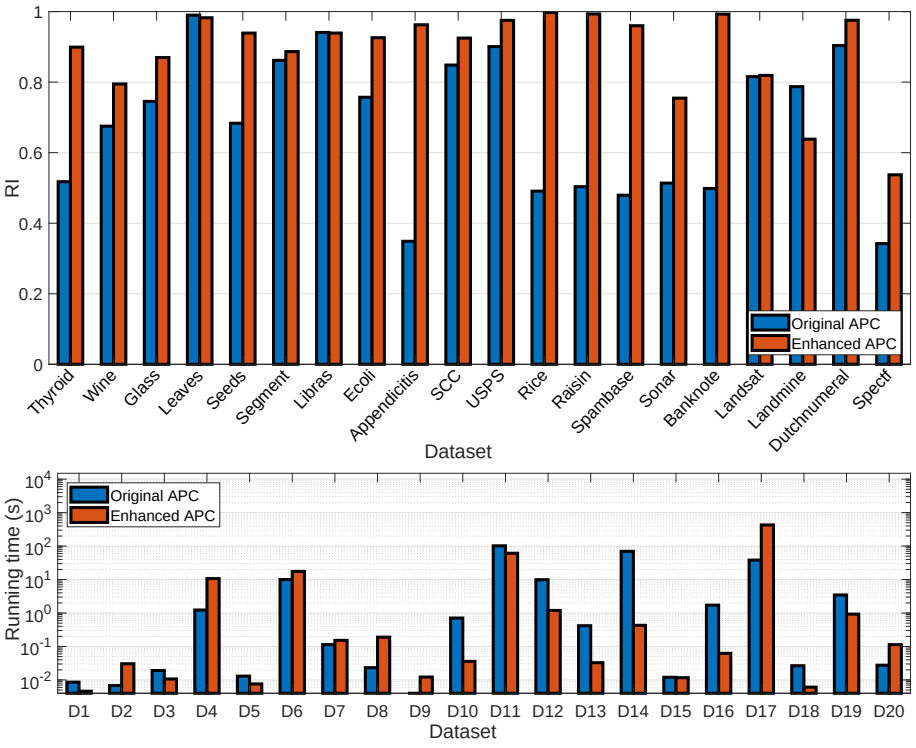

Figure 7: Comparison between the enhanced and original APC algorithme. The enhanced version generates better results with less running time on the majority of datasets.

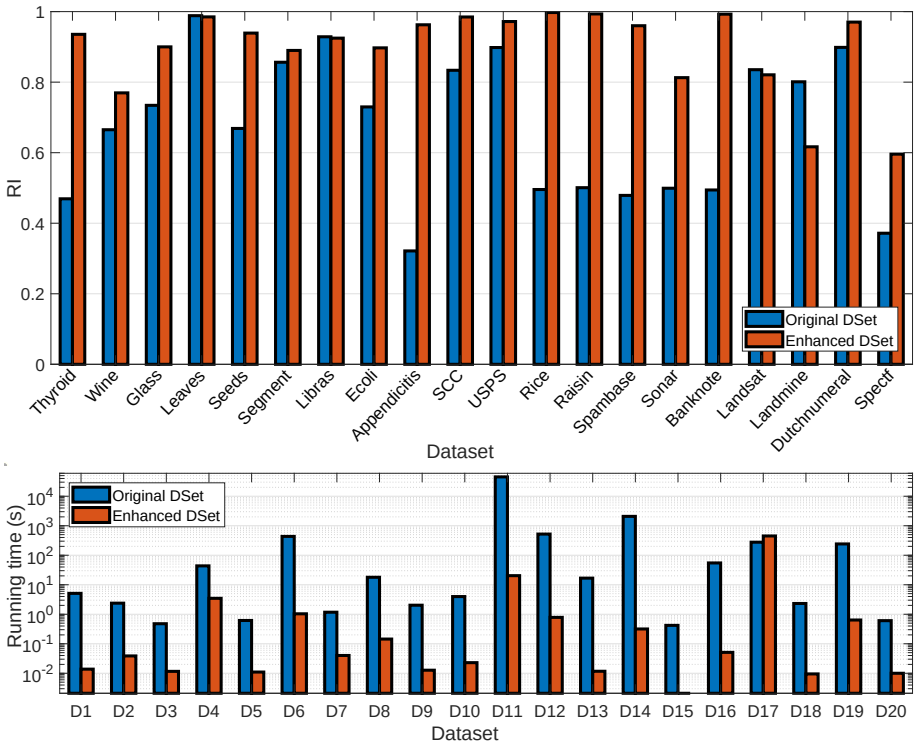

Figure 8: Comparison between the enhanced and original DSet algorithm. The enhanced version generates better results with less running time on the majority of datasets.

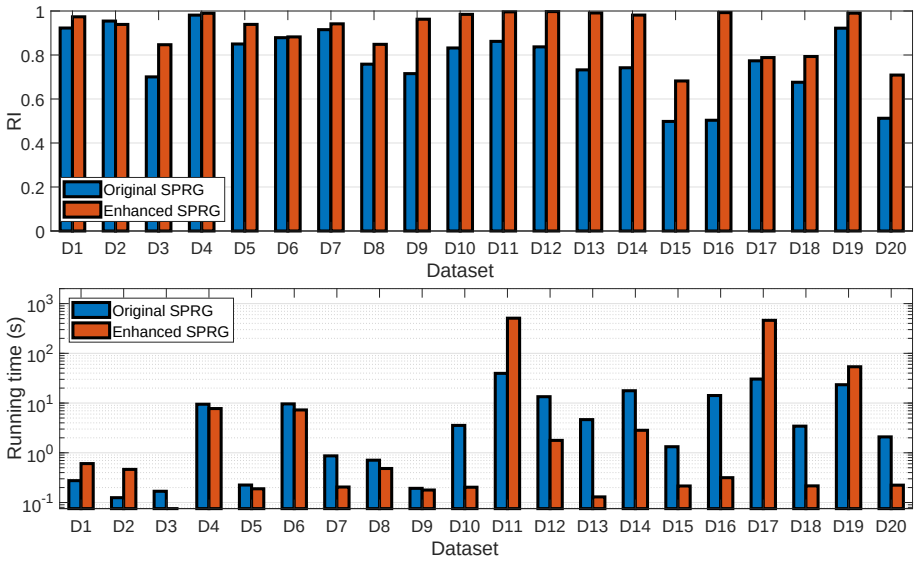

Figure 9: Comparison between the enhanced and original SPRG algorithm. The enhanced version generates better results with less running time on the majority of datasets.

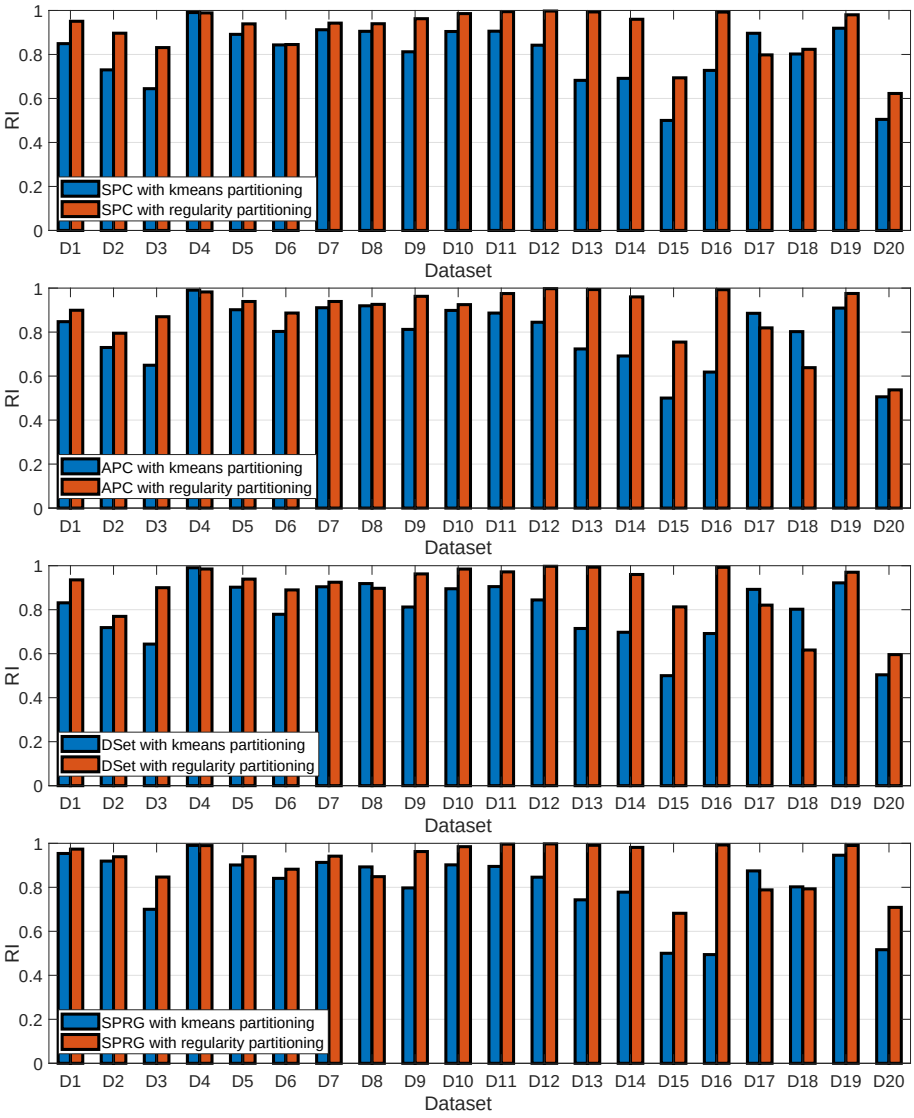

Figure 10: Comparison of RI between the enhanced algorithms with kmeans-based partitioning and with regularity partitioning. In most cases the regularity partitioning is shown to perform better than kmeans-based partitioning.