SUPPLEMENTARY MATERIAL: CLUSTERING OF MULTIPLE MICROARRAY EXPERIMENTS USING INFORMATION INTEGRATION

Elena Kostadinova[1*], Veselka Boeva[1], Niklas Lavesson[2]

[1] Computer Systems and Technologies Department, Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
[2] School of Computing, Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden
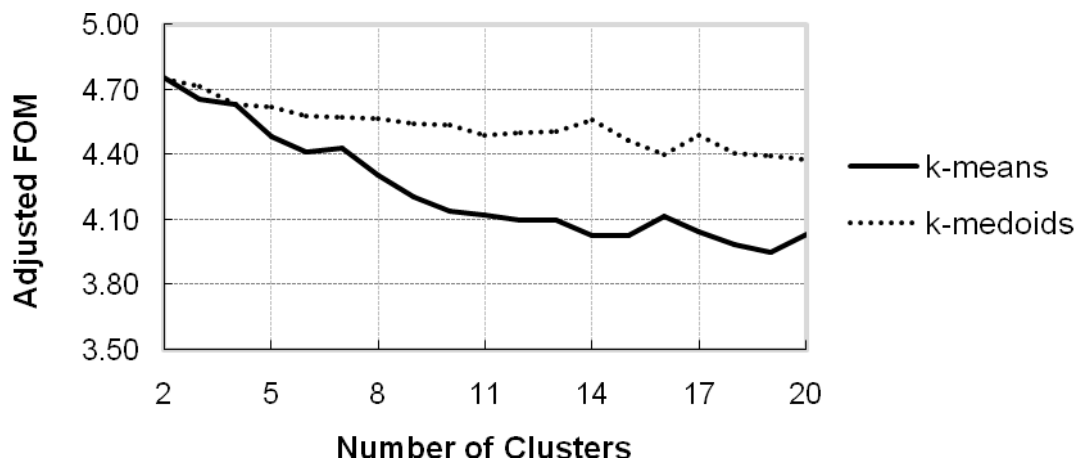[*] to whom correspondence should be addressed: elli@tu-plovdiv.bg

Figure 1: *Adjusted FOM* curves of *k*-means and *k*-medoids clustering methods for different number of clusters generated on the integrated distance (fused expression, respectively) *cdc* Rustici *et al.* datasets.
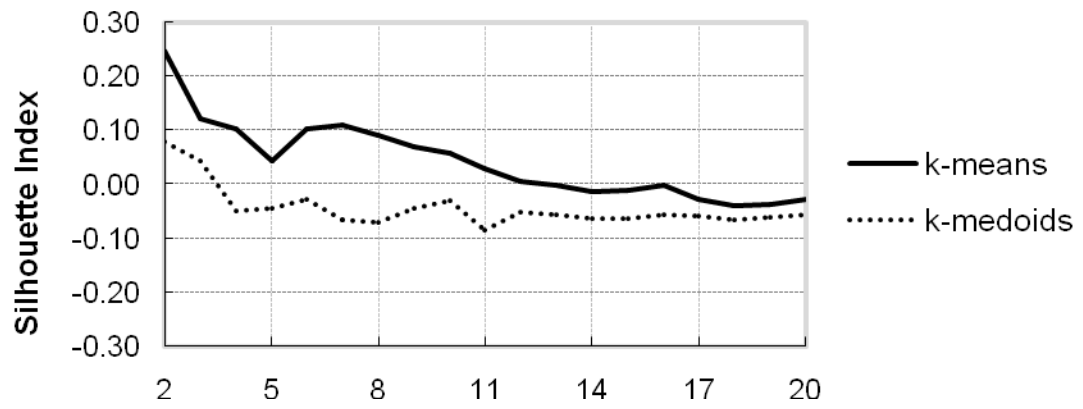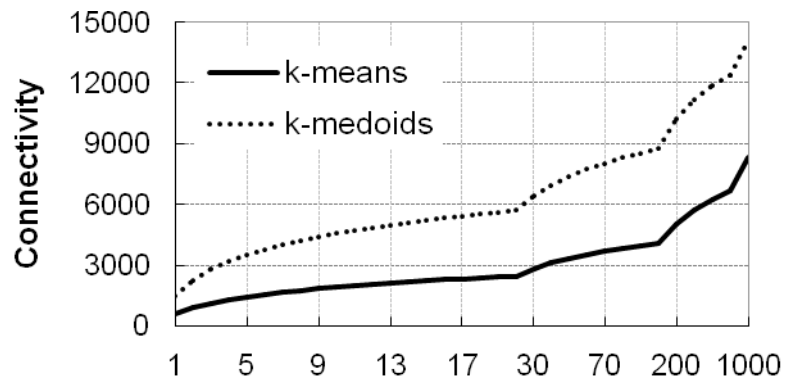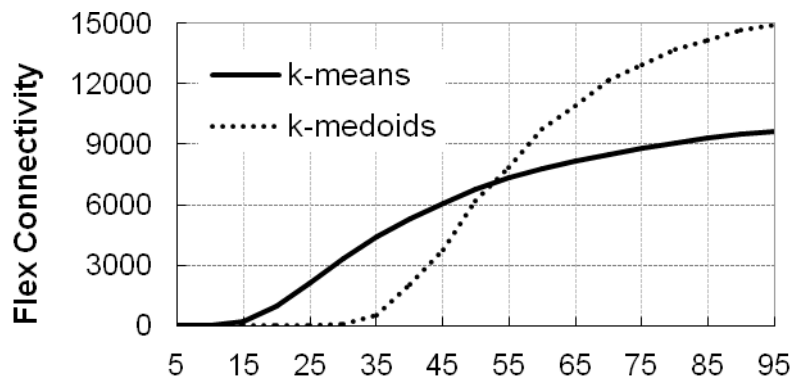
Figure 2: *SI* values generated by *k*-means and *k*-medoids clustering methods on the integrated distance and fused expression matrices of Rustici *et al.* datasets for different numbers of clusters.

Figure 3: *Connectivity* results generated on the combined Rustici *et al.* data for a range of different neighboring genes (a) and similarity cut-offs (b) using $k = 10$ clusters.
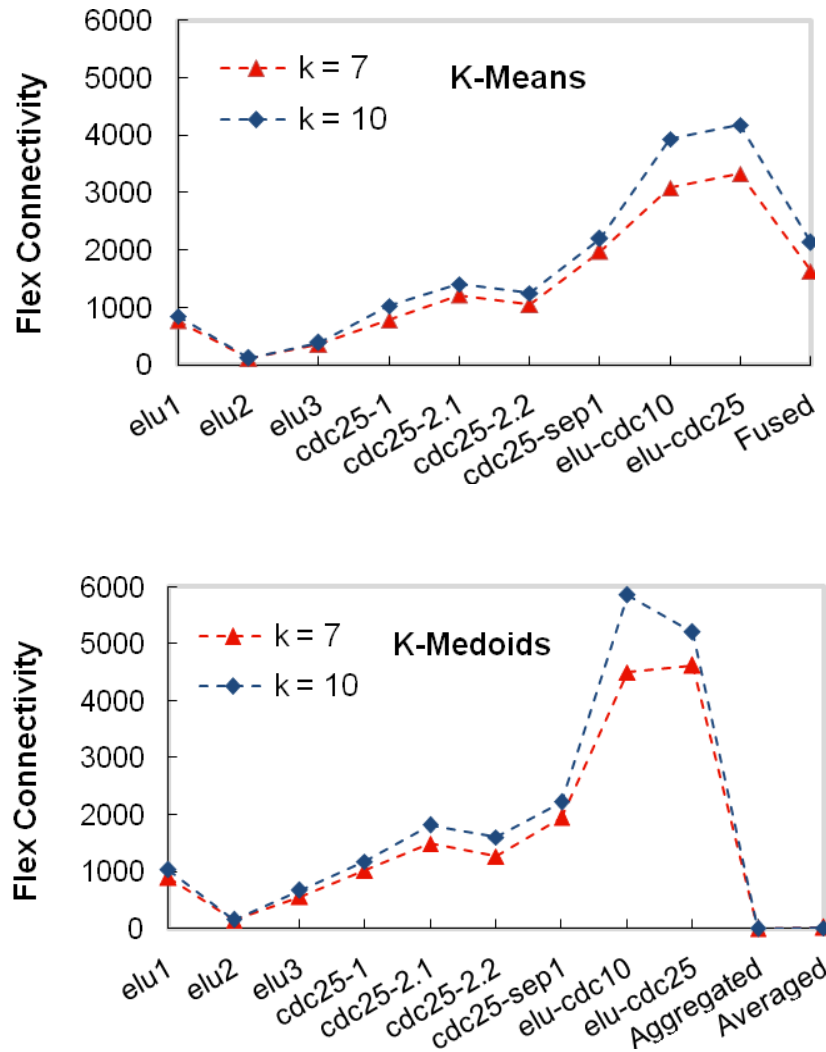
Figure 4: *Connectivity* results generated on the individual and combined Rustici *et al.* matrices for a similarity cut-off equal to 25% of the average distance.
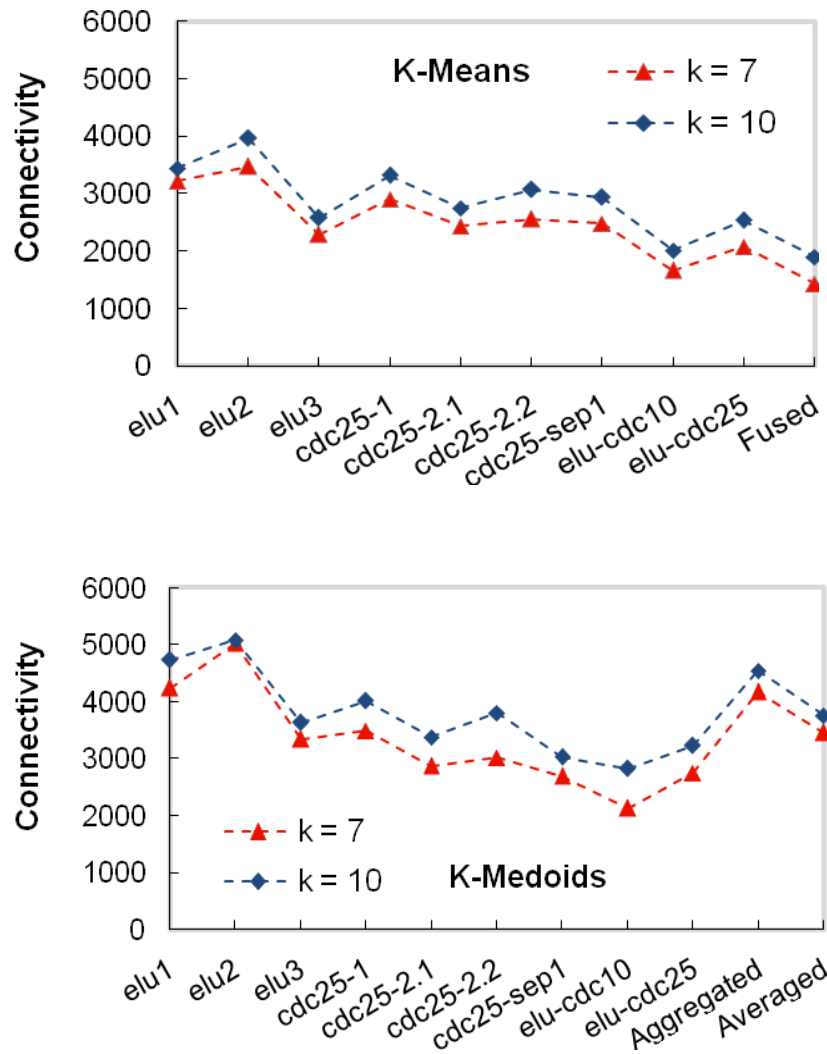
Figure 5: *Connectivity* results obtained on individual Rustici *et al.* datasets versus the corresponding values generated on the fused expression (integrated distance, respectively) matrix using a preliminary fixed number of neighboring genes (10).
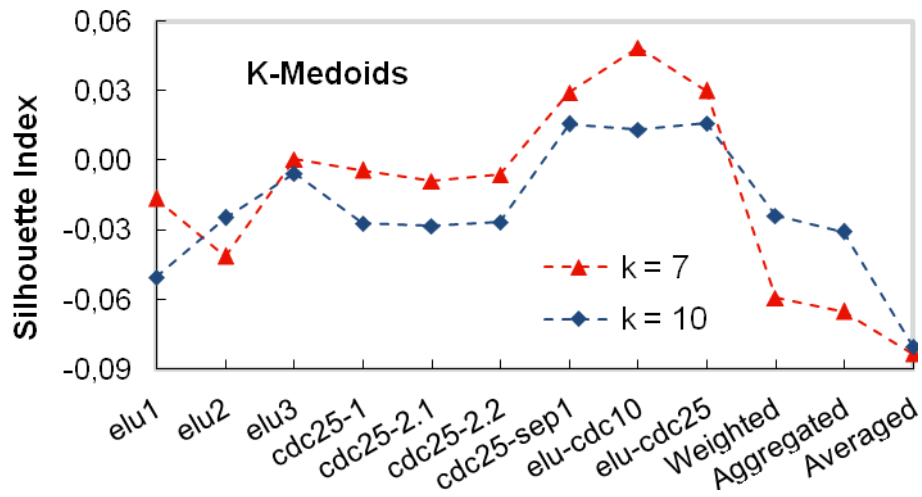
Figure 6: *SI* values generated on the individual Rustici datasets versus corresponding scores obtained on the weighted, integrated and the averaged distance matrices, respectively.
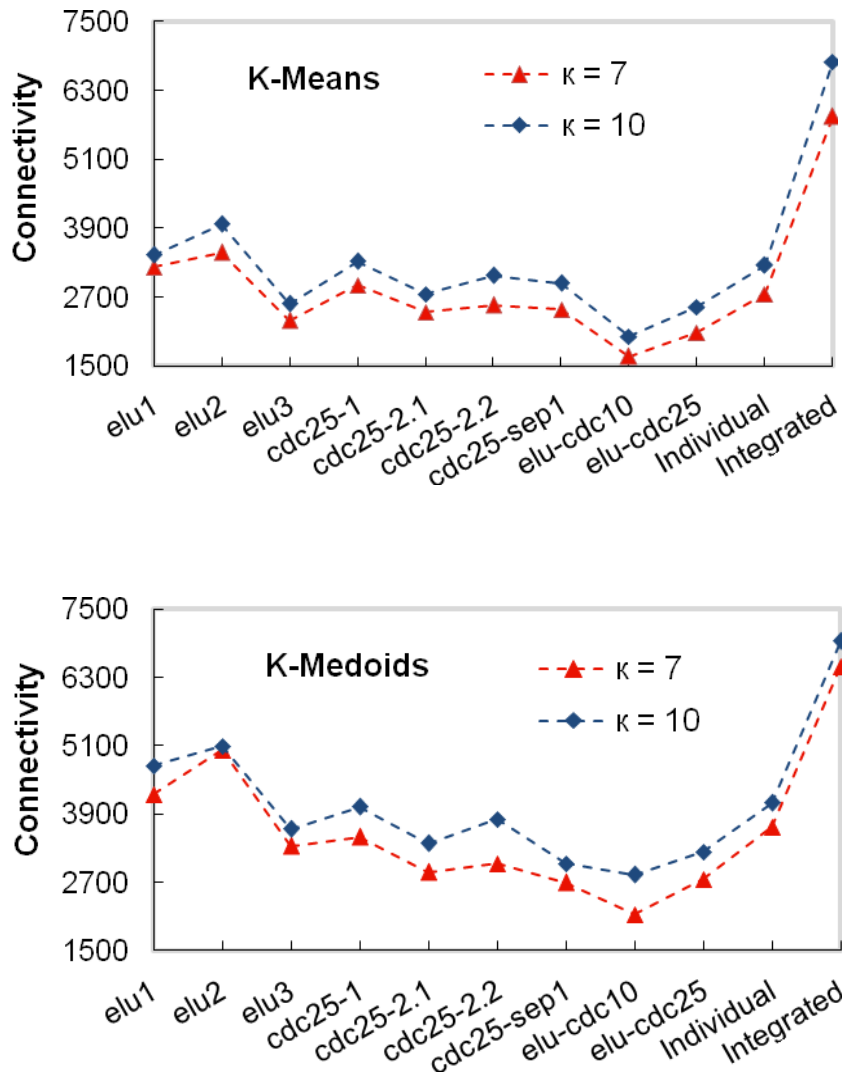
Figure 7: *Connectivity* scores generated on the individual Rustici *et al.* datasets versus means of connectivity values obtained on individual and integrated partitions, respectively.

Number of Clusters	k-means, %	k-medoids, %
$k = 4$	23.82	25.74
$k = 7$	31.73	34.05
$k = 10$	42.84	45.20
$k = 20$	47.10	46.70

Table 1: Average percentages of ties calculated for the both partitioning algorithms by using different cluster numbers.