

Supporting online material for:

Analysis of a Gibbs sampler method for model
based clustering of gene expression data

Anagha Joshi, Yves Van de Peer, Tom Michoel

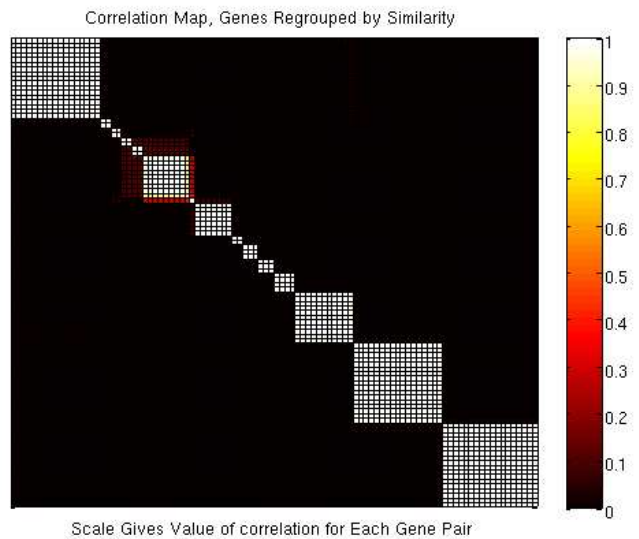
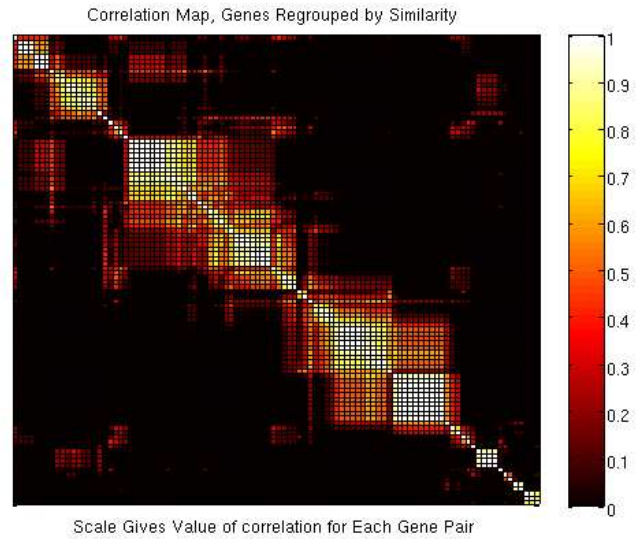


Figure 1: Gene-gene probability matrix of fuzzy clusters (above) and hard clusters (below) from 100 gene subset for the Gasch *et al* data set. Hard clusters are formed as a result of one run of Gibbs sampler while fuzzy clusters are formed by merging hard solutions of 10 different runs

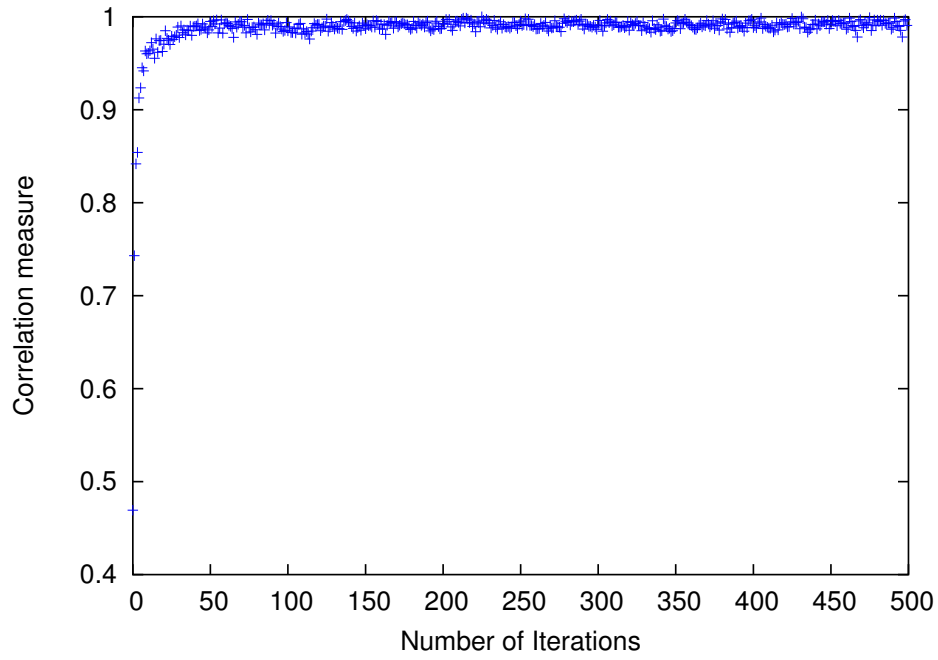


Figure 2: Correlation measure between two consecutive iterations for 1000 genes, 150 conditions dataset (subset of the Gasch *et al* data set) during 1 run. As it reaches almost one, it shows that algorithm gets stuck in local maximum.

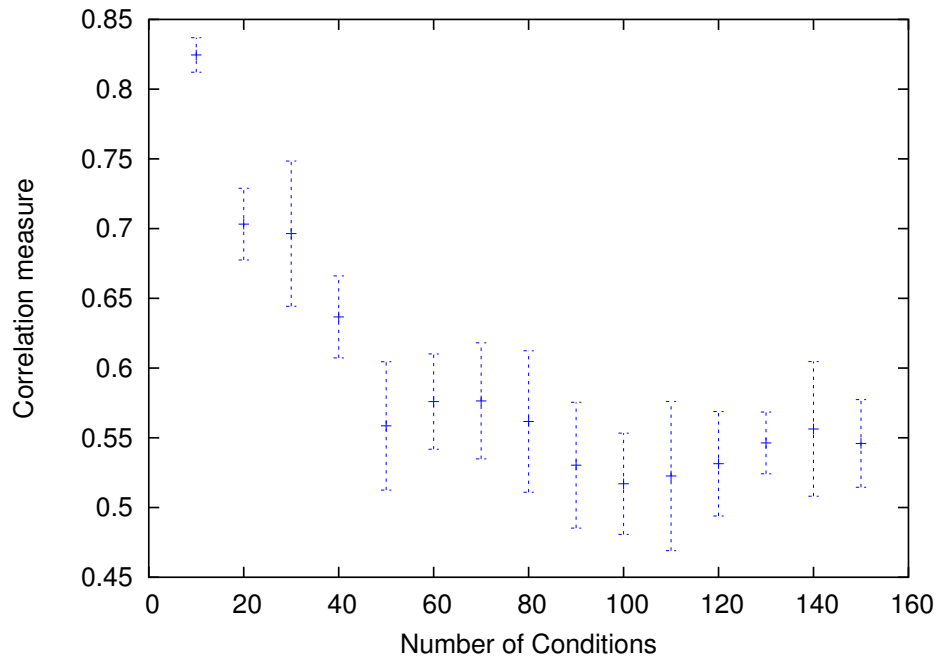


Figure 3: Correlation measure for data sets with 1000 genes and varying number of conditions (subsets of the Gasch *et al* data set)

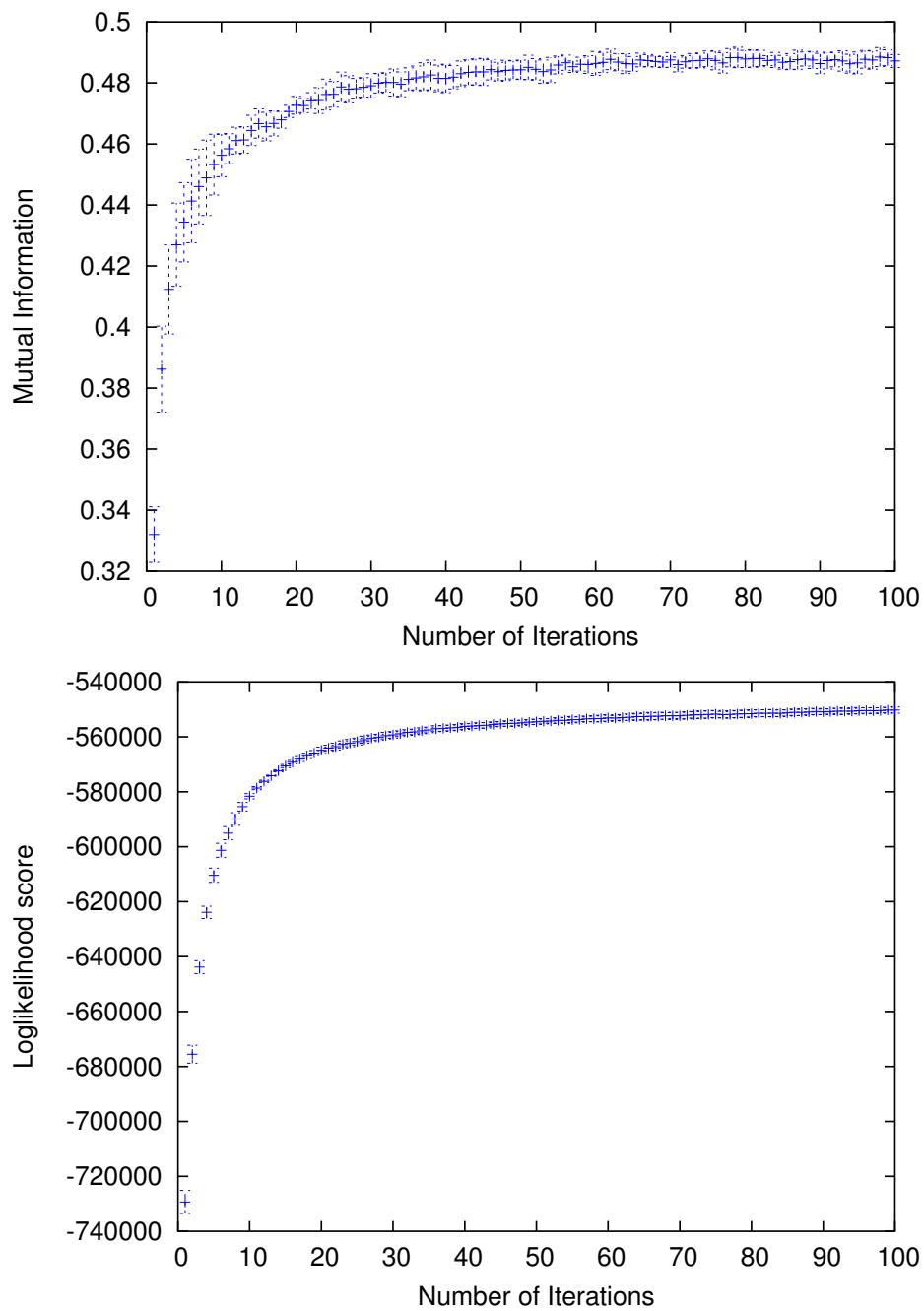


Figure 4: Average mutual information (above) and average log-likelihood score (below) for 10 Gibbs sampler runs for the Gasch *et al* data set.

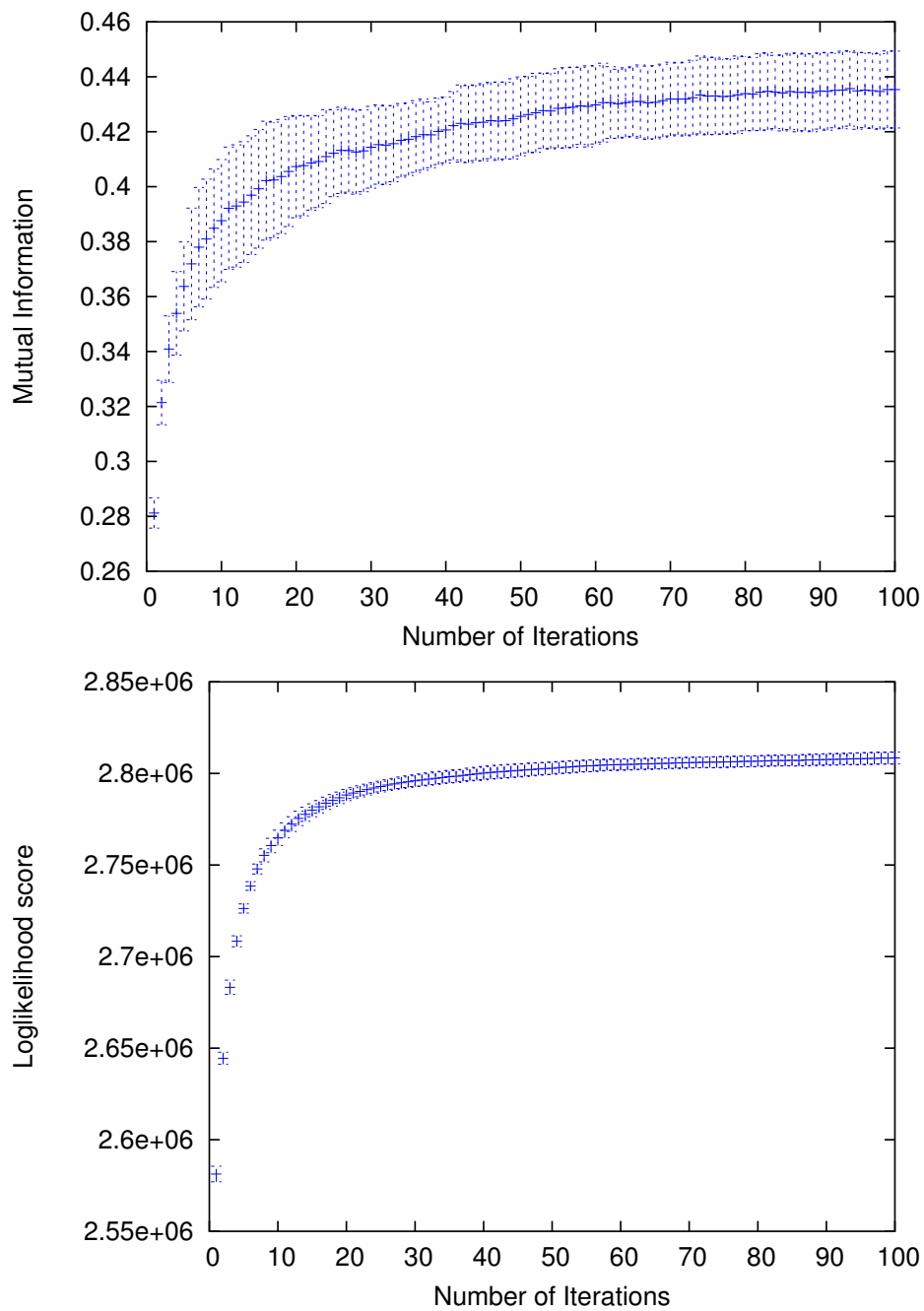


Figure 5: Average mutual information (above) and average log-likelihood score (below) for 10 Gibbs sampler runs for the Hughes *et al* data set.

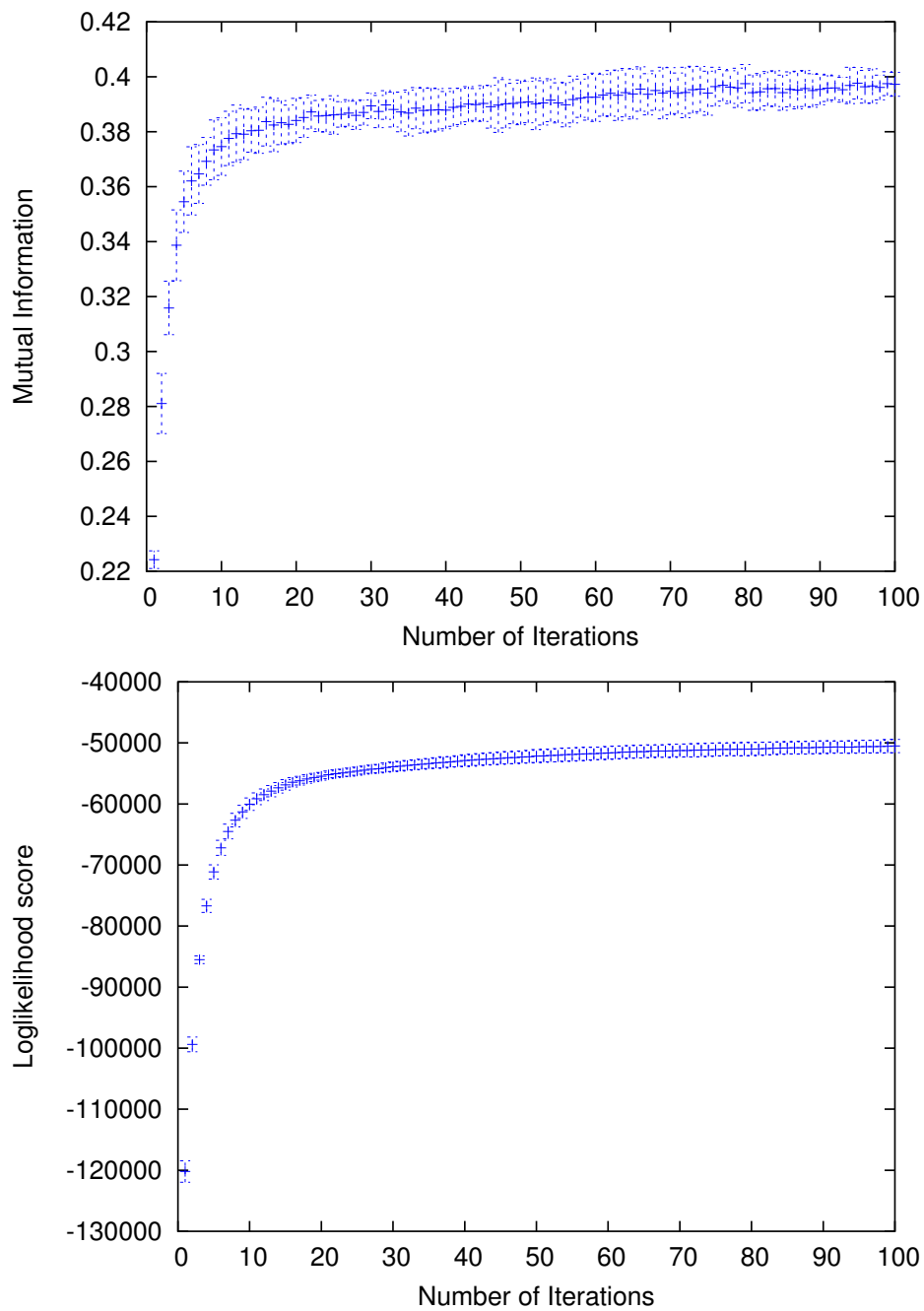


Figure 6: Average mutual information (above) and average log-likelihood score (below) for 10 Gibbs sampler runs for the Spellman *et al* data set.

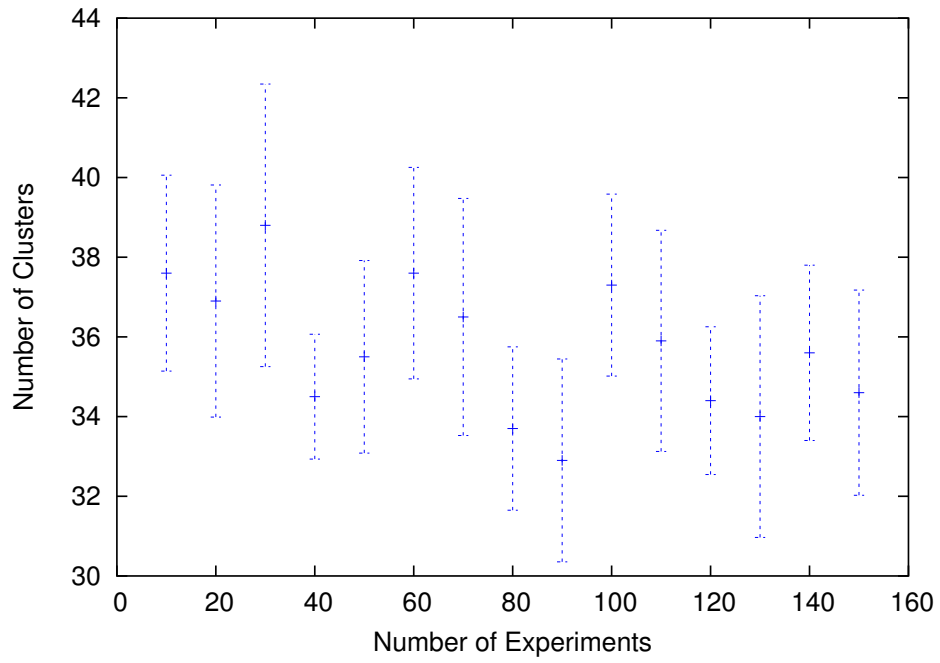


Figure 7: Number of clusters generated for data sets with 1000 genes and varying number of conditions (subsets of the Gasch *et al* data set). This shows that unlike simulated datasets, eventhough more experiments are added, the algorithm does not generate more number of clusters