1. Supplementary Material

Parameter Settings

The parameters and values used to generate results are provided in Table 1. We now show evaluations of our algorithm with various parameter values.

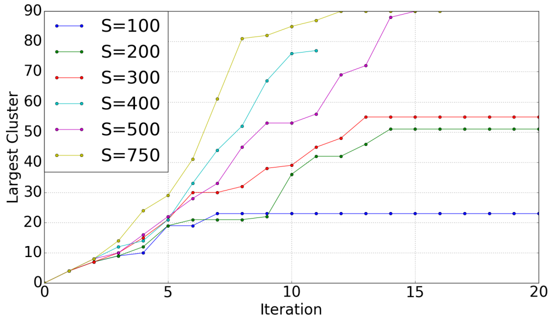

Figure 1: Cluster size growth rate at different number of clusters (S) selected during each iteration. Results for $S = 100$ and $S = 400$ converged, and were stopped earlier than 20 iterations. The plot shows the evolution of the largest cluster at every iteration, computed using different values of parameter S. This plot suggests that selecting 500-750 clusters each iteration yields best convergence.

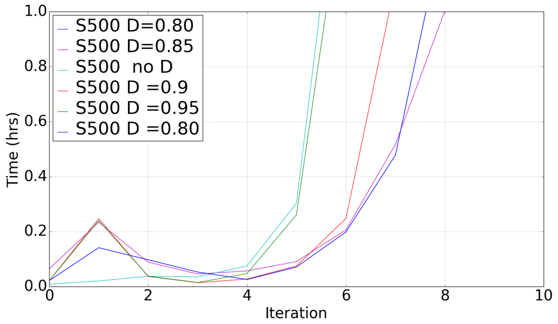

Figure 2: Comparison of the impact of diversification (selection) parameter on time. This parameter specifies the maximal Jaccard coefficient diversity in the population of clusters after any given iteration. It can be seen that the higher the diversity parameter (corresponding to more similar clusters kept), or when diversification is not performed, yields slower run time, at the expense of the quality of the reconstructed clusters.

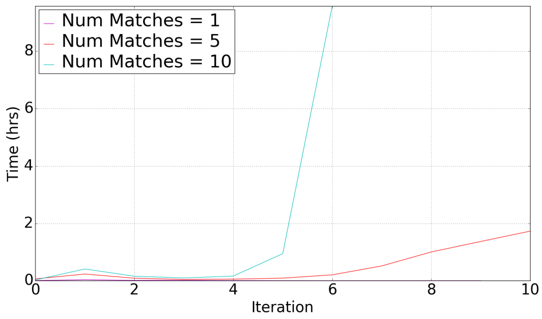

Figure 3: Comparison of the number of matches considered on time.This parameter controls the number of spanning matches selected from the randomized Monte Carlo match selection step to create a new cluster via match merge. Considering more matches yields a better F-score (corresponding to a larger and more accurate reconstruction), at the expense of time.

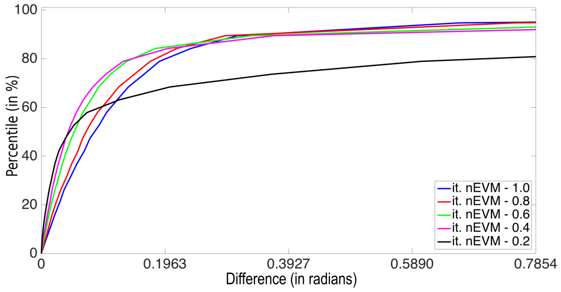

Figure 4: Comparison of different choices of match threshold (in radians) in the iterative Normalized Eigenvector Method (it. NEVM), and correct match recovery cumulative error (on a control dataset with fixed amount of noise added). As can be seen, parameters 0.4 and 0.6 yield best reconstructions, after which performance rapidly degrades.

Parameter	Description	Value
S	Maximum number of selected clusters in each iteration	500
Diversification	Maximum value of Jaccard coefficient allowed in selection step	0.85
W	Weight of spanning matches/fragments on fitness function	0.05
$\beta_{i,j}$	Multiplicative factor that influences the merge score based on number of added matches M after optimization	$1+0.1M$
V	Maximum allowable overlap proportion	0.65
C	Maximum allowable number of spanning fragments/matches in a cluster of N fragments	$N/50$
T	Threshold to discard matches in it. NEVM	0.6 or $.95 \cdot$ Error

Table 1: Parameter choices used for results displayed.

Mode	Largest Cluster (Fr.)	Match F-Score
Our Full System	90	0.823
Match-Only Merge	36	0.349
Fragment-Only Merge	17	0.204

Table 2: Comparison of merging strategies for reconstruction. Any one type of merge is insufficient to produce full reconstructions of the data, while the combination of both types of merges produces a large reconstruction.

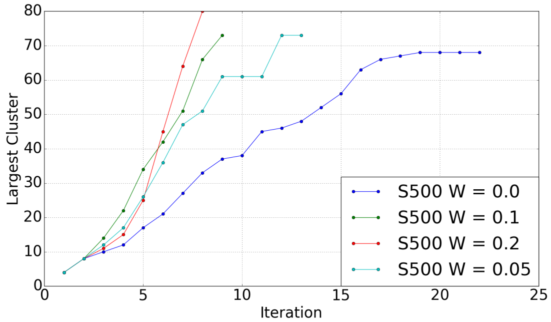

Figure 5: Influence of Spanning Match/Fragment Quantity (W) on the Fitness Function: Comparison of Reconstructed Cluster Size (run for maximum of 7 hours).

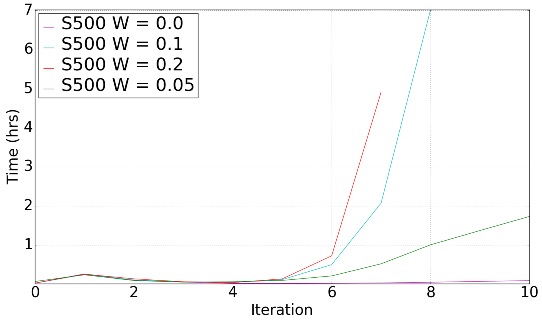

Figure 6: Influence of Spanning Match/Fragment Quantity on the Fitness Function(run for maximum of 7 hours): Comparison of Time Taken per Iteration. To avoid the significant increase of time associated with larger values, we used $W = 0.05$.