Which blind tract clustering method is most robust to false positives?

Mark Drakesmith[1], John Evans[1], Anthony David[2], and Derek Jones[1]

[1]*CUBRIC, Cardiff University, Cardiff, Wales, United Kingdom,* [2]*Institute of Psychiatry, Kings College London, London, United Kingdom*

Introduction Automated clustering of tractography streamlines into distinct bundles corresponding to known fasciculi is an important step in large scale analyses of tractography data and is thus an active area of research. Blind (as opposed to model / atlas based) clustering methods attempt to cluster tracts using parameters derived from the data. Previous studies have compared tract clustering methods[1] but issue that has not been adequately addressed is robustness to the presence of false positive (FP) reconstructions, which are more prevalent when using fibre orientation distribution (FOD) methods that resolve crossing fibres such as spherical deconvolution methods, than when using DT-MRI–based approaches. Ambiguity in the fibre orientation can result in streamline trajectories that do not correspond to the true anatomy. FPs will inevitably lead to inaccuracies when using blind clustering. Here, we systematically evaluate the robustness of some of the most widely-used blind clustering methods to the presence of FPs and characterize the type of misclassifications that are likely to occur.

Methods DWI-data were collected over 60 isotropic directions, and following motion and eddy current distortion correction, the FOD reconstructed in each voxel with the dampened Lucy-Richardson algorithm [2]. Streamlines were launched from every voxel in the 128x128x60 data set, and a subset of six pre-defined intra-hemispheric bundles was defined through manual segmentation (Fig. 1a), which we refer to as 'idealised'. A random proportion of streamlines in the idealised dataset were replaced by an equal number of randomly selected FP streamlines that did not conform to any of these 6 bundles (Fig 1b). The FP-rate, expressed as a fraction of the combined total number of streamlines across the six bundles, was varied between 0 and 0.5, (increment=0.1). 5 distance metrics calculated between all pairs of tracts (fig. 2). For each metric, k-means clustering was applied to: (i) the raw distance metric; (ii) an affinity metric computed by applying a Gaussian transform to the raw metric; or (iii) a set of spectral embedding vectors computed from the affinity matrix[5]. k=7 clusters were specified (6 true + 1 FP cluster). Performance was quantified as proportion of streamlines correctly assigned to a cluster (taken from the maximum across all permutations of cluster labels to accommodate arbitrariness of cluster labelling). This procedure was repeated 10 times, to quantify variance in performance.

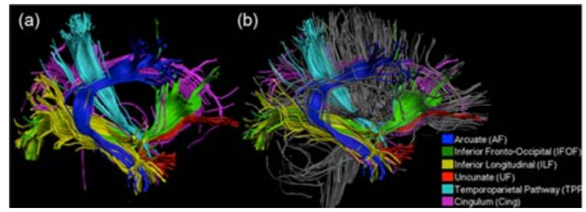

Fig 1 (a) Fibre bundles in the idealised data set (b) Idealised dataset with an FP rate of of 0.5 (FP tracts are in grey)

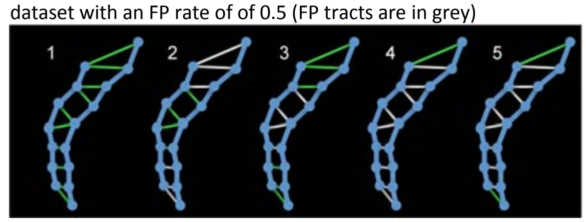

Fig 2. Distance metrics tested. (1) Mean of all points (Chamfer distance)[3]; (2) mean weighted towards middle of tracts; (3) mean weighted towards ends of tracts; (4) maximum (Hausdorff) distance[4]; (5) endpoint distance[5].

Results Fig. 3 shows metric performance *versus* FP-rate. All raw distance metrics, except the mid-weighted mean metric, show initial good performance but with gradual degradation with increasing FP-rate. Affinity-based distance metrics show similar responses but max and endpoint metric performed consistently high. Spectral embedding did not offer noticeable change in performance, although variability is higher. Fig. 4 characterizes the misclassifications, mean distance metrics have misclassifications favouring spatially adjacent bundles (e.g. AF→TPP and IFOF →ILF). The same pattern is seen for all raw distance metrics. For max and endpoint affinity metrics, misclassifications are more focal towards the smallest bundle (UF), while other misclassifications are minimal.

Discussion Distance metrics based on tract extremities rather than the middle body are most robust to FPs. Middle portions of tracts are not distinctive compared to endpoints. Computing the affinities further increases their distinguishability from noise. Spectral embedding does not offer functional improvement, but it is still preferable in order to reduce computational demands. Although it is beneficial for misclassification to be concentrated on a single bundle, the downsides it that smaller genuine bundles are less likely to cluster correctly.

Conclusions Our study shows that robustness to FPs is variable across different blind clustering algorithms. The most robust to FPs are endpoint and maximum (Hausdorff) affinity metrics. However, as we have shown, these methods are more likely to misclassify FPs with small bundles.

References **[1]** Moberts B, et al. *IEEE VIS.* 2005:65-72. **[2]** Dell'acqua F, et al. *NeuroImage*, 2010;49(2):1446–58. **[3]** Ding Z, et al. *Mag. Res. Med*, 2003;49(4):716–21. **[4]** O'Donnell L, & Westin, C-F. *MICCAI.* 2005:140–7. **[5]** Brun A, et al. *MICCAI.* 2004: 368–375. **[6]** O'Donnell L J, et al. *Am. J. Neuroradiol.* 2006;27(5):1032–6.

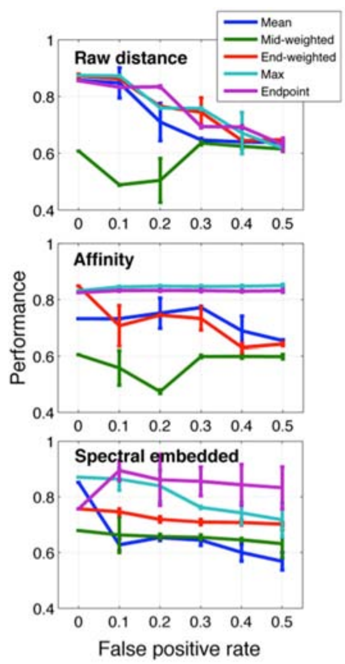

Fig 3. Mean and standard error of clustering performance for each clustering method across all FP rates.

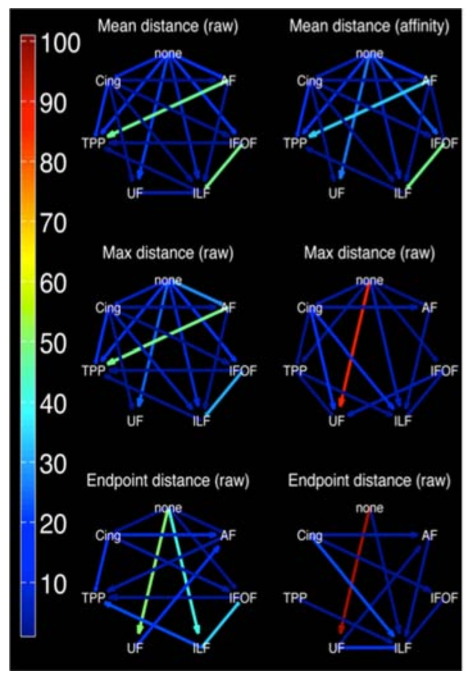

Fig 4. Mean percentage cluster misclassification for selected distance metrics. Colour-coded arrows show the percentage of streamlines being misclassified from one cluster to another.