# Meeting 2019/13/11

*Stephen Coleman*

*13/11/2019*

## Summary

Discussed possible projects for the first year.

### Projects

1. Consensus clustering;
2. Thesis introduction chapter;
3. TAGM-MDI (continuation of internship with Paul Kirk and Olly Crooke); and
4. Improving mixture MCMC by means of cluster-merge splitting.

For now I will work on Consensus clustering and the thesis introduciton. Current plan is to work on a 3/2 day rota for Consensus clustering (3 days) and the thesis introduction chapter (2 days).

### Consensus clustering

The aim is to turn the original project into a paper. To do this further case studies are required. I plan to use the following datasets:

1. Simulation where MDI does converge;
2. Simulation where MDI does not converge in a feasible time;
3. Subset of yeast genes used in original MDI paper;
4. Full set of yeast genes for which MDI cannot converge; and
5. A subset of the CEDAR dataset.

### Introduction chapter

The aim is to have an introduction to the thesis (focusing on integrative clustering methods). Current methods to include are:

1. MDI;
2. iCluster (Bayes);
3. Clusternomics;
4. Bayesian consensus clustering;
5. COCA;
6. JIVE; and
7. MOFA.

There exist other methods which are possibly of interest, but this is the initial set. I will compare these methods under some headings such as:

- Model description (for instance many are example of Dirichlet Multinomial Allocation models);
- Information sharing (how this mechanism works; arguably a subset of the above, but quite important);
- Software implementation; and
- Scalability.