Problems: Workshop on Topological Data Analysis and Clustering, April 22, 2022

The following is a distillation of the issues raised during the Problem Session, with attributions.

1) **Shiebler**: Given a multiparameter hierarchical clustering or persistent structure, how can we flatten the clustering efficiently?

2) **McInnes**: There should be a topological method for analyzing clusters directly from graphs (like UMAP) in a high dimensional setting, without resorting to dimension reduction (layout). What is it?

See McInnes' lecture and notebook.

3) **Healy, McInnes**: Is there a good topological interpretation of modularity for the graphs in Problem 2)?

4) **Scoccola**: How does one extract a single clustering from a multiparemeter heirarchical clustering? (main question of his talk)

5) **McInnes**: Adamyk presented a combinatorial theory of layers, and Scoccola presented a linearized approach, both for clustering in a multipersistence setting. Is there a construction that amalgamates these two approaches?

5.5) **Jardine** Given the endgame of Scoccola' talk, Problem 5) seems like the comparison of the algebraic K-theory of a field with the algebraic K-theory of the category of finite sets (aka. the stable homotopy groups of spheres). There should be a comparison map (combnatorial to algebraic) relating these. Does any of this give algorithms?

6) **Weir**: Is there a meaningful distribution theory for persistent homology groups, for random choices of N points on a fixed compact manifold?

7) **Weir**: Can one "metrize" the UMAP graph?

8) **Scoccola**: Is there a coarsening of the UMAP graph that would give a good approximation?

9) **Jardine**: The construction of the UMAP graph uses probabilistic choices of k-nearest neighbours. How does varying these choices affect the topological structure of the filtered graph?