**SDA Final Project**

**Individual differences emerge from behavioral data**

**Based on:** Forkosh, O., Karamihalev, S., Roeh, S. *et al.* Identity domains capture individual differences from across the behavioral repertoire. *Nat Neurosci* **22,**2023–2028 (2019). <https://doi.org/10.1038/s41593-019-0516-y>

*Background, Motivation and Goals*

Current main methods for measuring individual differences have several very broad and unnatural assumptions, accordingly their results scope is limited since they mostly relay on questionnaires. subject to lab-based conditions and derive from small-scale datasets. In contrast, study above (Forkosh 2019) offers an innovative noble computational scheme to assert the behavioral basis of individual differences from semi-naturalistic personality analysis on mice.

Under the umbrella of the SDA course, we challenge the data analysis part preform on this paper on synthetic humans behavioral data with similar features but with 5 “real” clusters (due to the notorious Big 5 personality traits model). We implement an exploratory analysis to test both the IDs and personality space they suggest. For that we plan to find (a) unique identity domains (IDs) and (b) unique roots (architypes) to these IDs so they form a personality “space”, which spans all the subjects personalities based on these IDs. In addition, we challenge the results from the paper to see (c) how sample size affects the results (on random data, but future work with real data is theoretically possible).

*Results*

1. Identity Domains (IDs) – we search for the best features to describe IDs of behavioral data. We implement different dimensionality reduction techniques under different optimization conditions (variability within groups vs. complete data) using different techniques - PCA, kernel PCA, our own LDA, ISOMAP and t-SNE.

* **Only the Manually implemented LDA managed to emerge the clusters from our random behavioral data.**

1. Personality Space – we search for the best model to describe archetypes of behavioral strategies exhibited in nature. For that we implement different clustering analysis techniques - Hierarchical clustering and ICA, to see how the subjects behavioral data spilt among clusters.

* **While the former differs given different sample sizes, the latter and LDA successfully divided the space into 5 clusters, yet a triad of 3 can fit as well.**

The code for my analysis and main figures were uploaded to GitHub ([link](https://github.com/YuvalSK/SDA-Final-Project) - figures in Results folder).