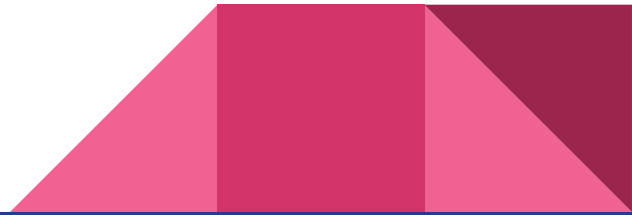
FFT-accelerated Interpolation-based t-SNE

Eric Roeck

The Software

Fast Fourier Transform - accelerated Interpolation-based t-distributed Stochastic Neighbor Embedding (FIt-SNE)

- An algorithm developed by the Kluger Lab at Yale Medicine
- Created as an improvement on Barnes-Hut t-SNE (BH t-SNE), the previously fastest t-SNE implementation
- Developed to reduce processing times for large datasets
    - At the Kluger Lab, it is used for visualizing single-cell RNA-sequencing data



Implementation of FIt-SNE

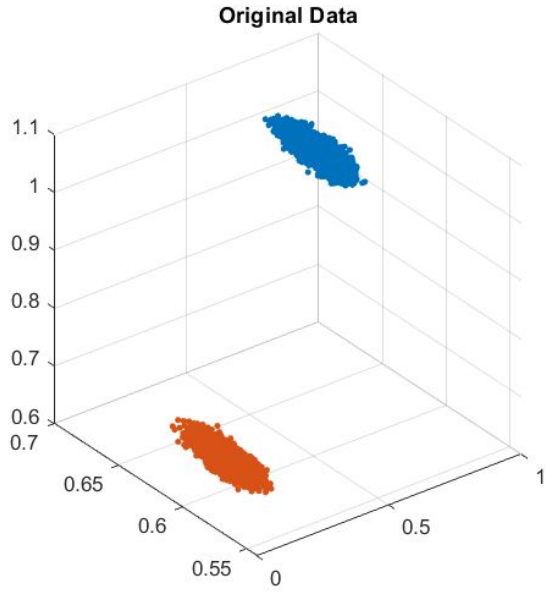
- The algorithm is significantly faster
  - O(N) vs O(N^2)
  - BH t-SNE uses approximation, FIt-SNE uses interpolation
- The code is written in C++ and can be run on R, Python, and MATLAB wrappers provided on git

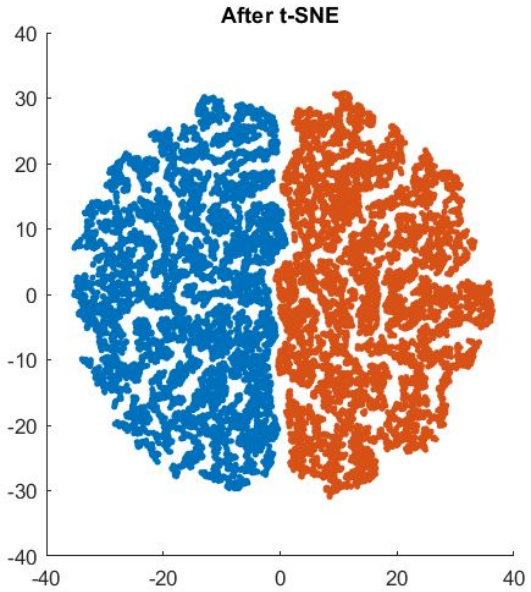
Table 1 | Time taken for 1,000 iterations of the gradient descent phase of 2D t-SNE using BH t-SNE compared to our implementation (FIt-SNE), as compared on a 2017 Macbook Pro for a given number of points *N*

N	BH t-SNE	FIt-SNE
10,000	1 min	<1 min
100,000	11 min	<1 min
500,000	1 h 10 min	3 min
1,000,000	3 h 9 min	15 min

See the Methods for more details.

Some Data





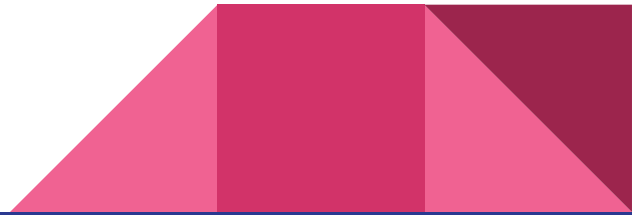
I’m not a biologist I don’t know what this means but it looks neat right?

Future Work

- Verify performance of FIt-SNE and BH t-SNE
- Compare the accuracy the two approaches

Why do we care?

- t-distributed Stochastic Neighbor Embedding has many uses in data science
    - Neural networks, genomics, computer security, music analysis, etc.
- It would be interesting to learn R to run it



Links

Github:
https://github.com/KlugerLab/FIt-SNE

Journal:
Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data