CPSC 453 Problem Set 2: Graph Clustering and Signal Processing

Author: Wenxin Xu

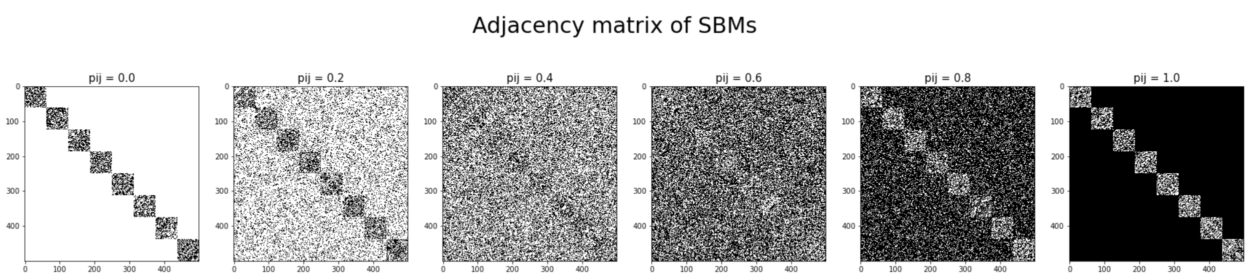
The `xu_wenxin_ps2.zip` file includes 3 files:

- `xu_wenxin_ps2_report.pdf` : A detailed report.
- `/code`
  - `ps2_functions.py`: contains 9 functions (`load_json_files, gaussian_kernel, sbm, L, compute_fourier_basis, gft, filterbank_matrix, kmeans_plusplus, SC`)
  - `xu_wenxin_ps2.ipynb`: A Jupyter Notebook contains all the code.

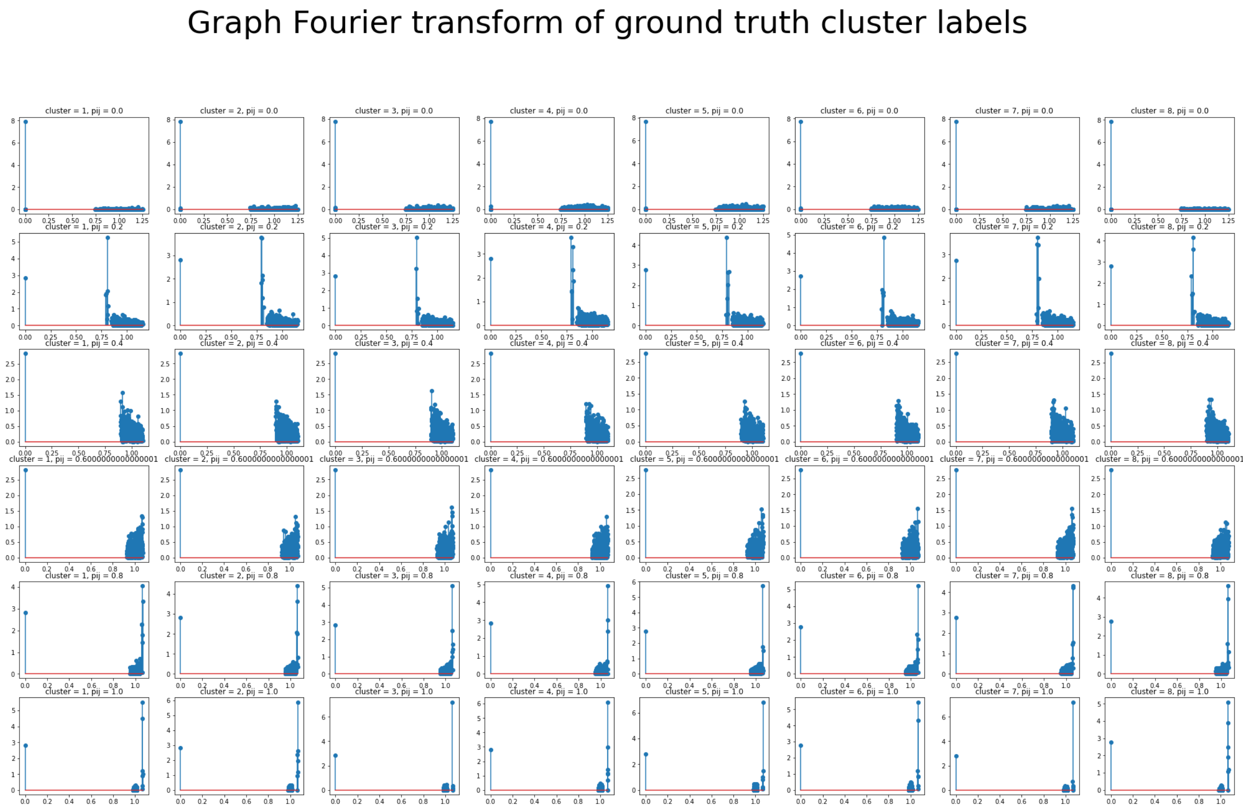
3 Filtering signals on stochastic block model

Generate a set of SBMs with various values of p_{ij}. Start with $p_{ij} = 0$ and work in small increments until the graph is highly connected between clusters. Use 500 points and 8 clusters.

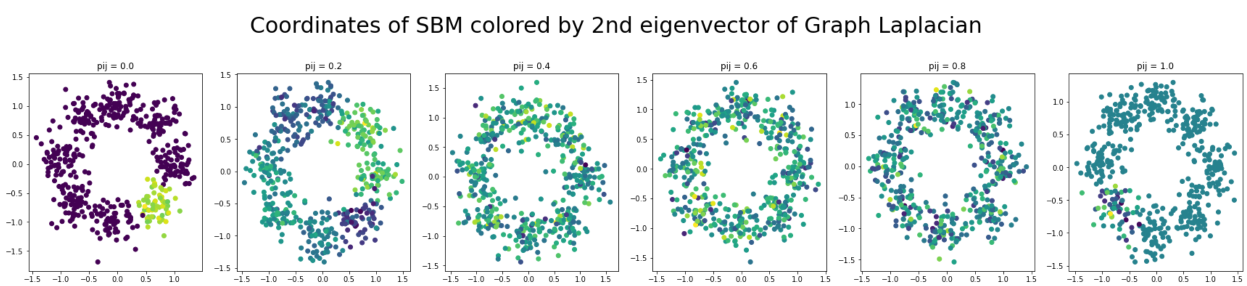
1. Visualize the adjacency matrix of your SBMs



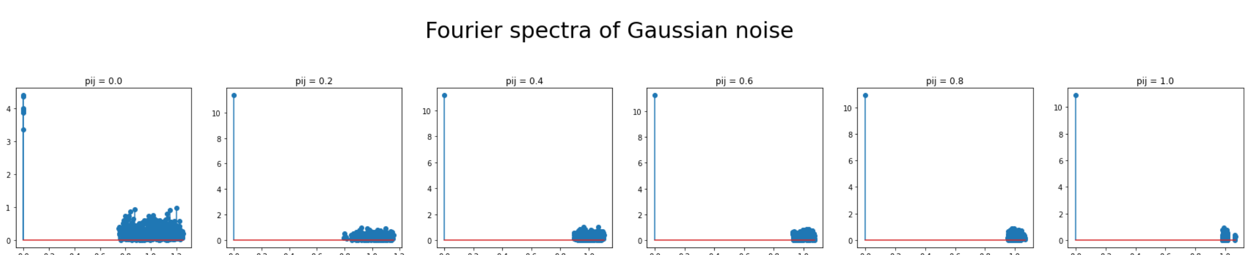
2. Plot the absolute value of the Fourier transform of your ground truth cluster labels for each SBM realization.

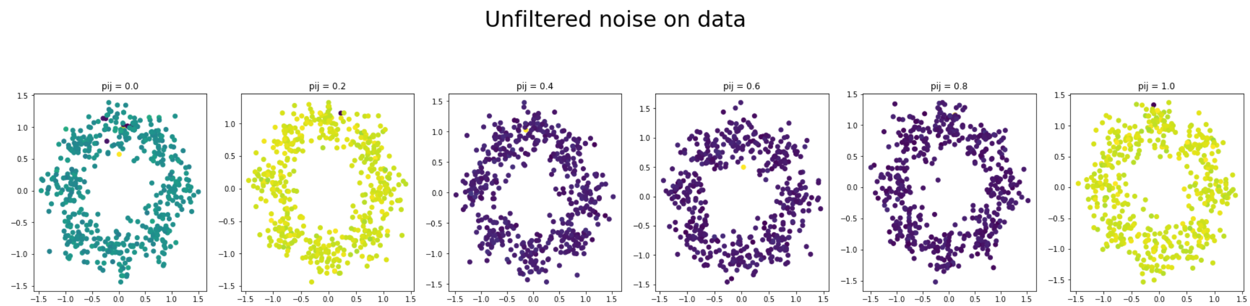


3. At each realization, plot the coordinates of the SBM colored by the second eigenvector of the graph Laplacian.

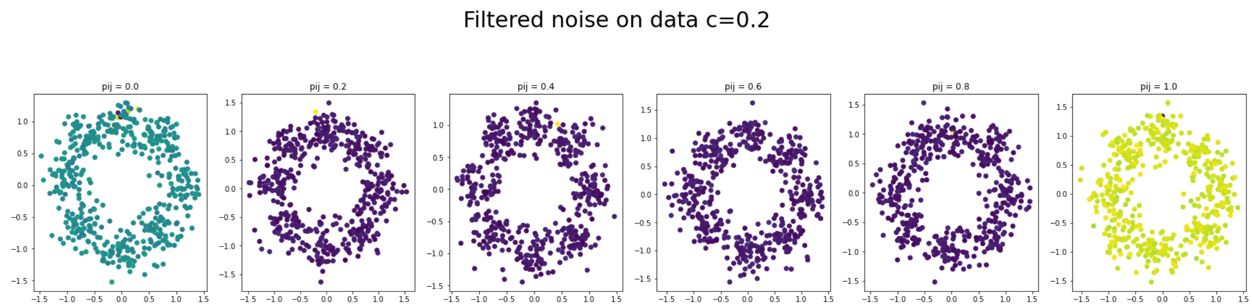


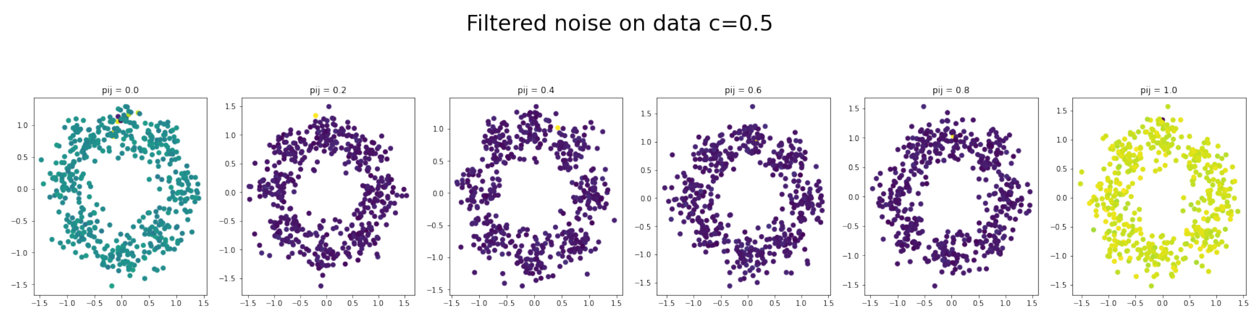
4. Use your filter function to filter random Gaussian noise. Try low pass filters which are "ideal", that is, if $\lambda > c$, the filter evaluates to 1. Otherwise, the filter evaluates to 0. Plot the Fourier spectra of a few realizations of Gaussian noise, and plot the filtered noise on the data coordinates.

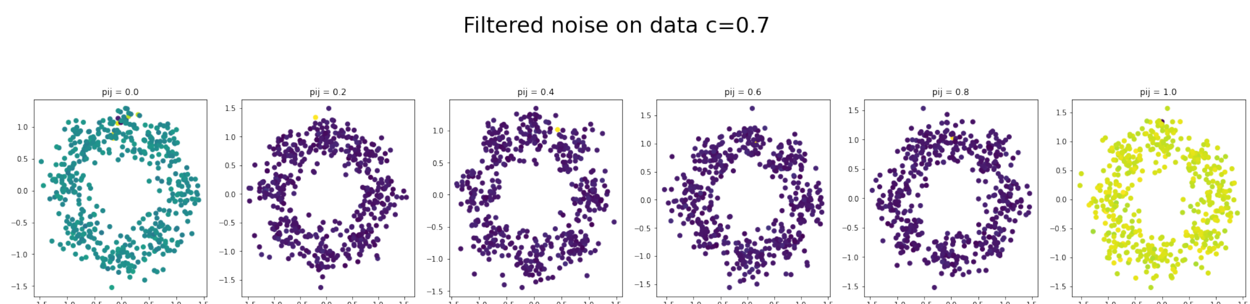




I try different threshold c for the low pass filter (c = 0.2, 0.5, 0.7). The results are similar.







Question 3.1. *How does the spectrum of the cluster labels changing at various levels of connectedness? What kind of frequency content does a stable clustering have?*

I start pij from 0, and increases pij by 0.2 until 1 and set pii = 0.5. The spectrum of cluster labels is the distribution of graph Laplacian eigenvalues of Graph Fourier Transform of cluster labels, represents the frequency content of signal (i.e., cluster labels). When pij = 0, there is no connection between clusters, only connections within clusters, so the frequency content of signal is low. As pij increases, the connection between clusters is increasing, as we can see the plots become more and more dense, means that the frequency content of signal becomes higher.

Question 3.2. *What do the first few Laplacian eigenvectors represent in the SBM? Why?*

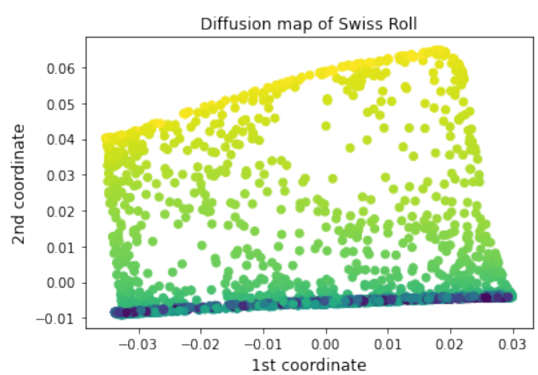
In SBM, the first few Laplacian eigenvectors represent low frequency because they are correspond to small eigenvalues. The first non-trivial eigenvector correspond to an eigenvalue (Fiedler value) which represents the minimum cut needed to split the graph. The third eigenvector and the subsequent eigenvectors after the third can discriminate between clusters while the top 2 eigenvectors don't.

Question 3.3. *What does the filtered noise look like on average? What happens when you vary the parameter c?*

The unfiltered noise on data looks like full of noise while the filtered noise on data looks uniform among clusters. So the filter works. I first start with threshold c = 0.5 for my filter, then I tried c=0.2, 0.7. The higher the threshold, the more signals remaining after filtering.

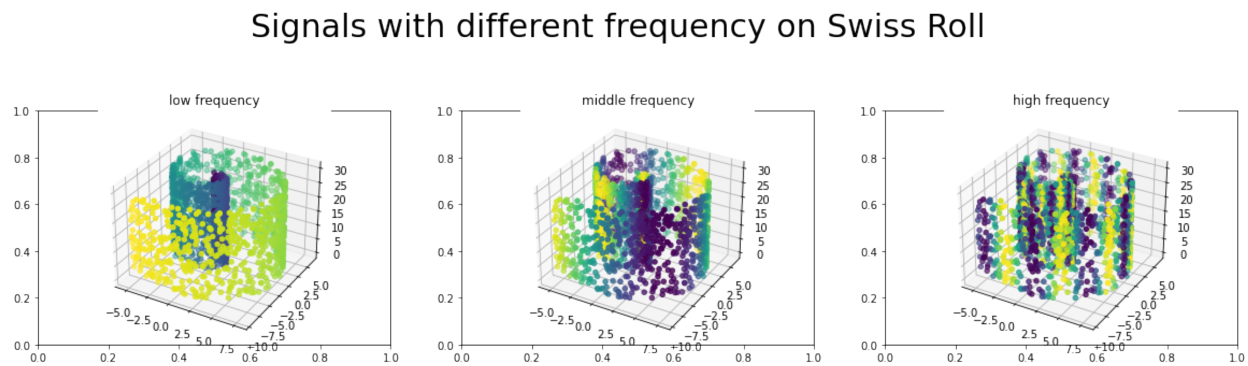
4 Filtering Signals on the Swiss Roll

Using the Swiss roll from Assignment 1, build a graph using the Gaussian kernel parameters that produce a reasonable diffusion map (i.e. it looks like a plane). Use this diffusion map to visualize the signals you generate below.



1. Use the coloring from the previous problem set to generate 3 signals: one low frequency, one medium frequency, one high frequency. Plot these signals and show their graph Fourier transform. Report the function that you used to generate them.

I use sin(1), sin(20), sin(100) to generate signals of low, medium high frequency, respectively.



GFT of signals with low frequency:

```
[ 1.69119744e+01 -1.68134744e+00  1.30243754e+01 ...  3.70550264e-03
 -2.99460235e-02  1.61456909e-02]
```

GFT of signals with middle frequency:

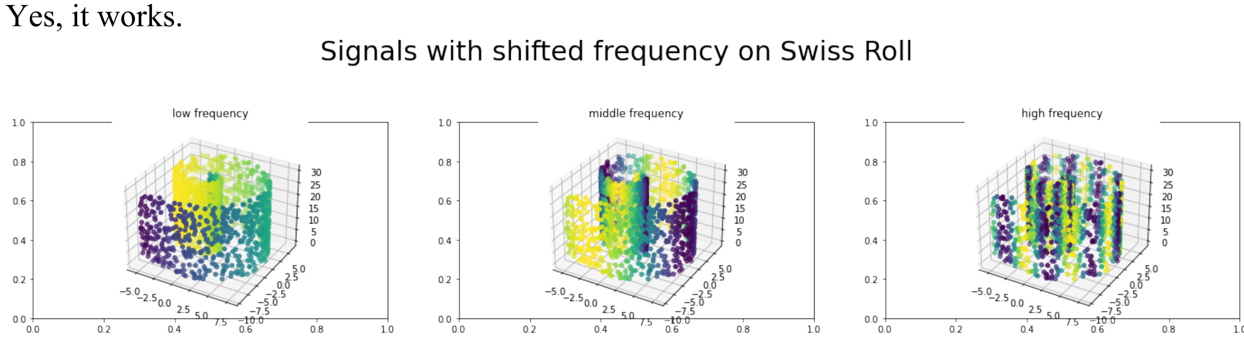
```
[ 2.06970375  2.12722276 -4.3806663  ...  0.01415984 -0.0354968
  0.01974694]
```

GFT of signals with high frequency:

```
[-0.36621316  1.60228947 -0.43904356 ...  0.02105947 -0.17205006
  0.15423249]
```

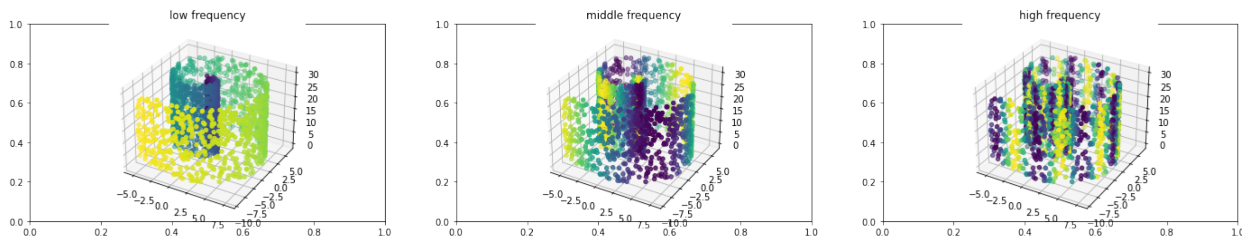
2. Try shifting the phase of your signals. Does it work?

Yes, it works.

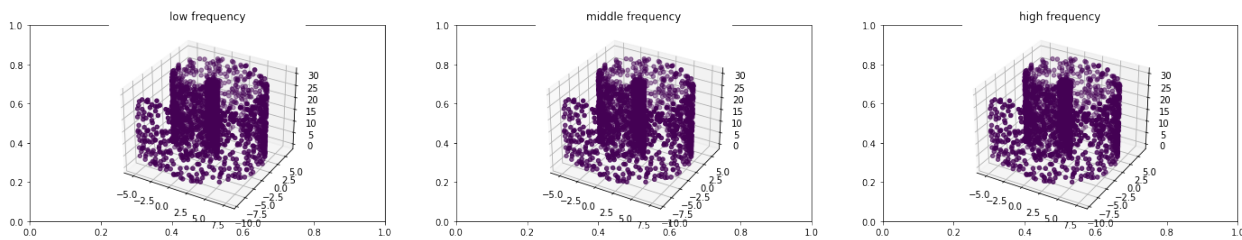


3. Use your filter function to filter these signals. Try ideal low pass filters and ideal high pass filter. Plot the result of two of your favorite filters for your three signals.

Signals with Low pass filter on Swiss Roll



Signals with High pass filter on Swiss Roll

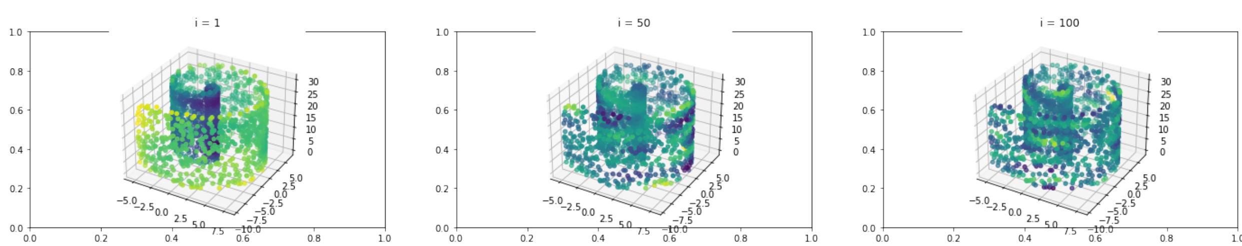


4. Craft a band pass filter using a Gaussian function. You can change the mean μ of the Gaussian to a desired target λ_k in the middle of the band. You can tune the width by changing σ.

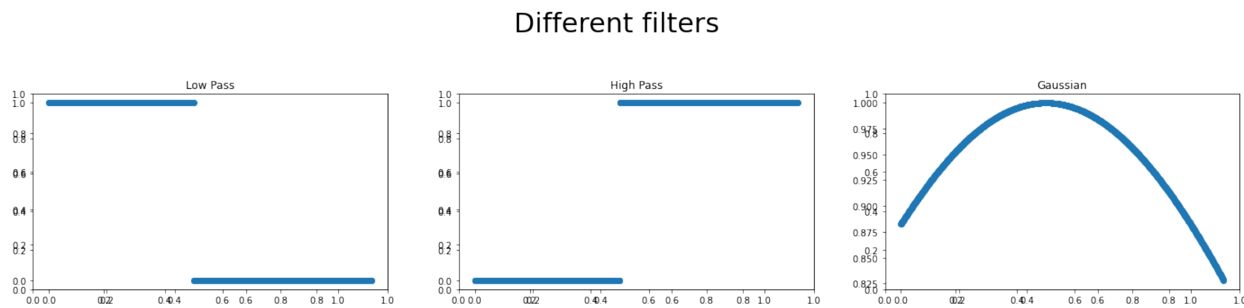
Use the Kronecker delta as a signal to translate your Gaussian to a certain point on the graph. You can do this by taking $H\delta_i$. Changing the value of i should move the Gaussian to different parts of the graph. Translate it over the Swiss roll and note what changing λ_k does to the output.

mean = 0.5, sigma = 1.0

Gabor filter on Swiss roll



5. Plot all of your filters by evaluating them over the interval $[0, \lambda_N)$.



Question 4.1. *What does smoothness mean in terms of graph signals and their frequency spectrum?*

There are several ways to measure smoothness: Laplacian quadratic form and graph spectral representation. In the graph signal domain (vertex domain), smoothness is related to connectivity of a graph. A smooth signal can't feature values that decay too quickly from the peak value. In the frequency spectral domain, the Graph Fourier Transform of a smooth signal decays rapidly. The smooth signal can be compressed because they can be approximated by just a few graph Fourier coefficients. A smooth signal has energy in low frequencies in the graph spectral plot, while a less smooth signal has energy in high frequencies.

Question 4.2. *Band-limiting is a nice trait if one wants to design an algorithm. It means that there isno frequency content above a certain frequency, i.e. the band limit. For what values of p_{ij} are SBM cluster labels band-limited?*

The band limit is the threshold value above which all the Graph Fourier Transform values are near 0. With pii = 0.5, when pij = 0, the band limit ≈ 0.7, when pij ∈ {0.2, 0.4, 0.6, 0.8}, band limit ≈ 1.2.

Question 4.3. *Under what scenarios would you want a band pass or high pass filter? How do these filters work and what do they do to the spectrum of the signals?*

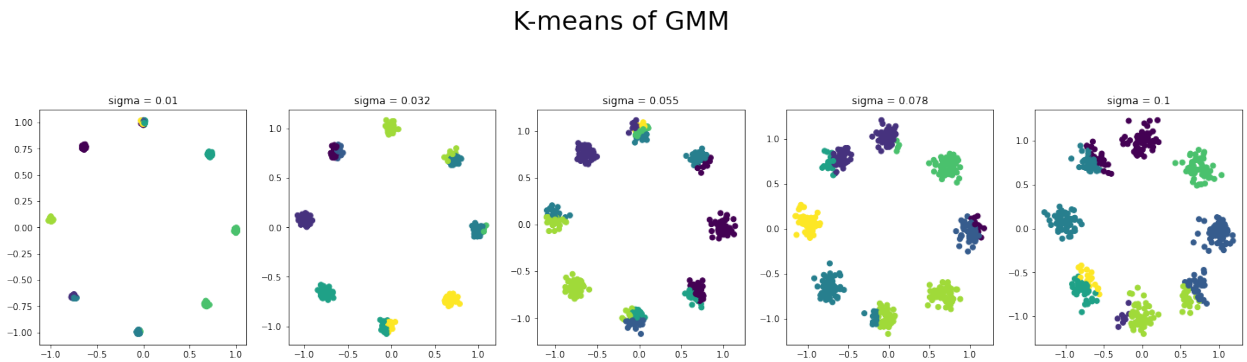
Band pass filter: i.e. Gaussian filter. If we want signals pass within a specific range of frequencies (frequency band), decay outside the range, also we want the transition between unfiltered signal and filtered signal is smooth, we can use a band pass filter. High pass filter: if we want high-frequency signals pass for a threshold while low-frequency signals decay, we can use a high-pass filter.

Question 4.4. *(Bonus) Are there any similarities between the band pass experiment and classical translationand modulation?*

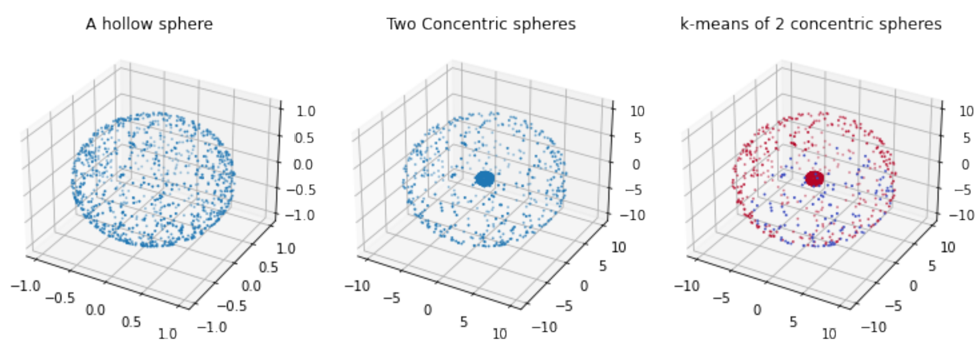
Yes. Band pass filter decays signals that doesn't follow the Gaussian distribution. Classical translation magnifies signals that follow the Gaussian distribution and keeps signals that doesn't follow the Gaussian distribution. Modulation keeps and magnifies all of the relevant signals.

5 k-means Clustering

1. Run k-means on your GMM with various values of sigma.



2. Next, generate a concentric spherical dataset by sampling 1000 points from a 3d unit normal distribution. Normalize all points using Euclidean distance. This should form a hollow sphere. Multiply the first 500 points by 10. This should create concentric spheres. Each sphere will be a separate cluster. Run k-means on this.



Question 5.1. *Compare the labels you obtain on the SBM to the ground truth clusters. At what values of sigma does k-means clustering start to fail?*

I test 6 values of sigma (0.01, 0.032, 0.055, 0.078, 0.1), when sigma = 0.032, the k-means clustering can't correctly differentiate all the ground truth clusters.

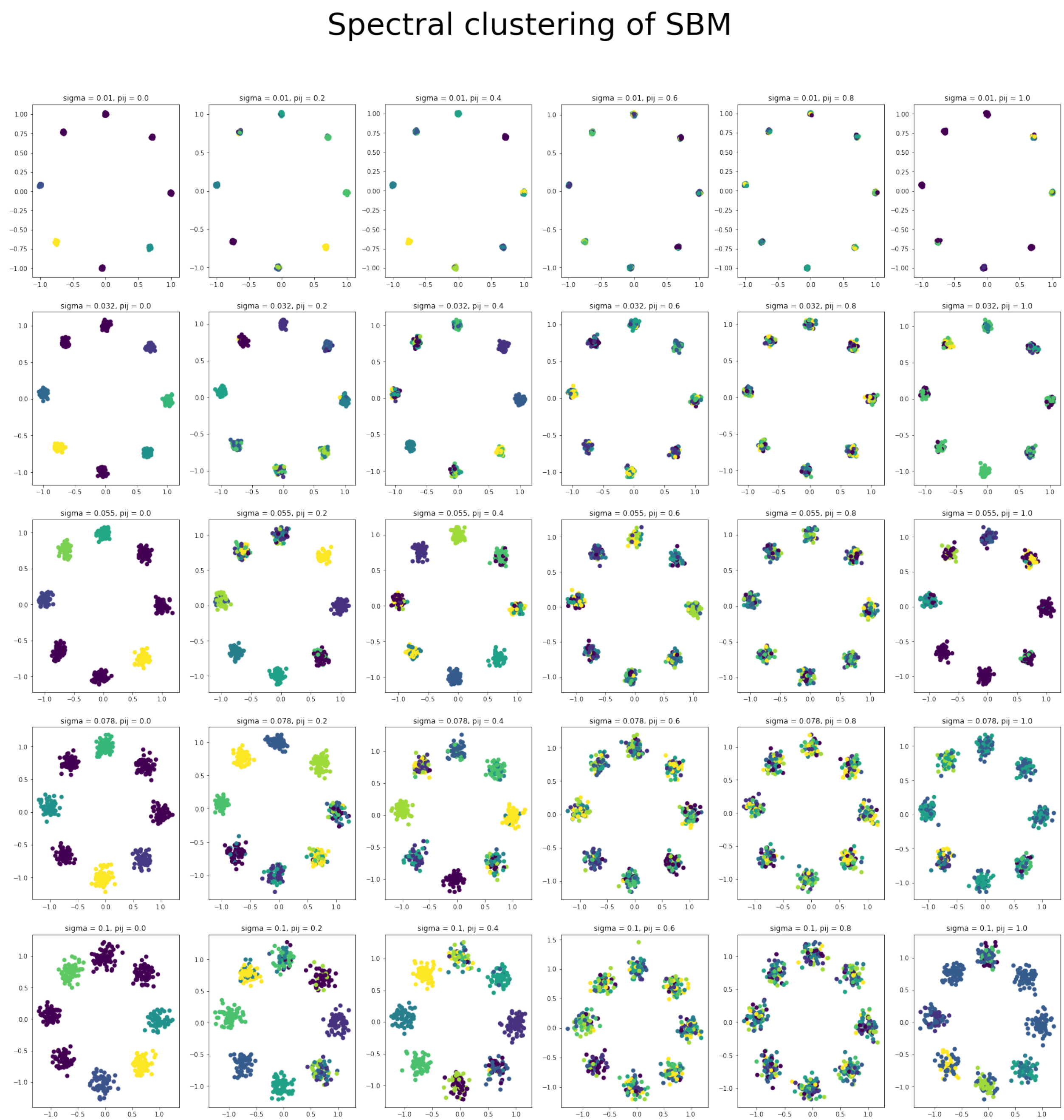
Question 5.2. *Does k-means converge to a reasonable clustering on the concentric spheres? Why?*

No. We can see in the k-means plot of 2 concentric spheres when using k=2, both the outer sphere and the dense inner small sphere are colored by red. This is because the k-means method can only distinguish between clusters with different means, while in this case the means of the 2 concentric spheres are the same: [0, 0, 0].

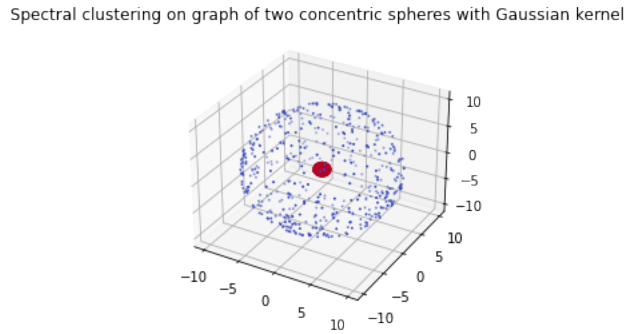
6 Spectral Clustering

Repeat our previous k-means experiment but using spectral clustering:

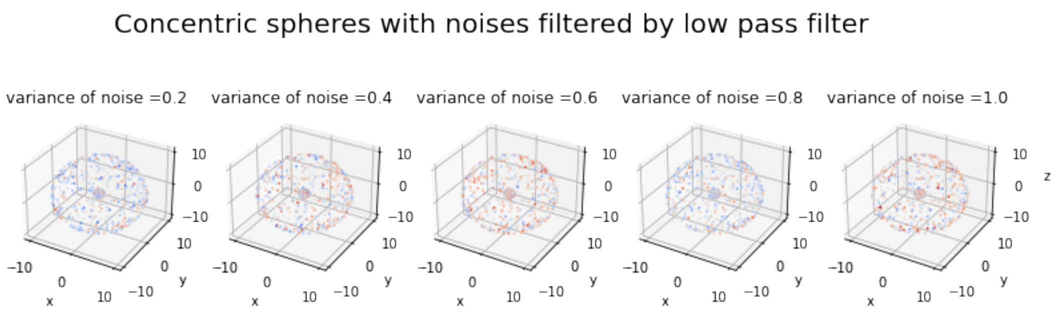
1. Run spectral clustering on your SBM with various values of sigma and pij.



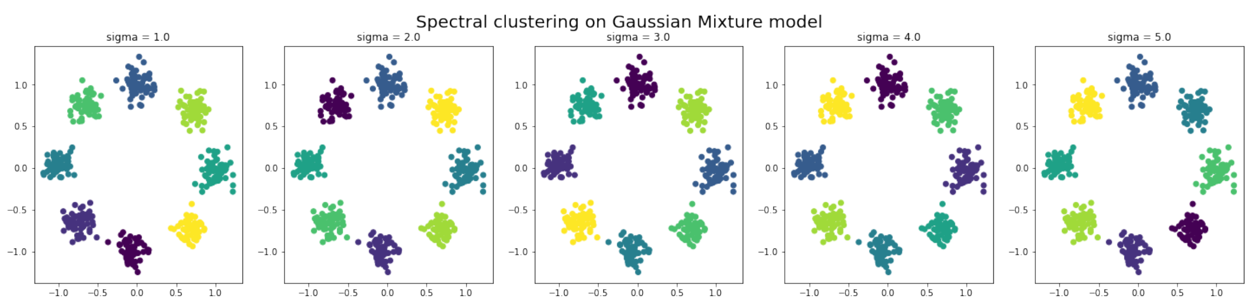
2. Next, Run spectral clustering on the concentric spherical dataset created in Section 5.



3. Generate multiple instantiations of Gaussian noise on the concentric spheres (vary the variance of the noise) and filter them using an ideal low-pass filter.



4. Finally, use your SBM to generate coordinates (as in regular k-means), but generate an adjacency matrix for these points using the Gaussian kernel (This concoction is called a Gaussian Mixture Model). Perform spectral clustering on this at various bandwidths.



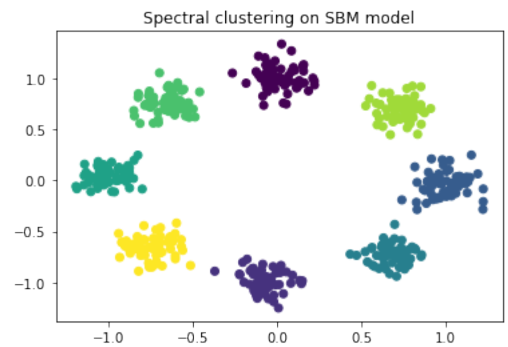
5. Compare these results to the ones you found in our initial k-means experiment.

Question 6.1. *Compare the labels you obtain on the SBM to the ground truth clusters. At what values of sigma and p_{ij} does spectral clustering start to fail?*

I try 30 different combinations of sigma from (0.01, 0.032, 0.055, 0.078, 0.1) and pij from (0, 0.2, 0.4, 0.6, 0.8, 1) to create SBM and run spectral clustering on them. In my case, no matter what values of sigma is, at pij = 0.02, the spectral clustering starts to fail, but as sigma increases, the degree of failure also increases.

Question 6.2. *How does spectral clustering compare to k-means on the SBM? How does using the SBM adjacency matrix compare to creating one from a Gaussian kernel?*

Spectral clustering is better than k-means on SBM. The figure below is spectral clustering on SBM without Gaussian kernel. Compared to the figure generated in Step 1, I don't see significant difference between using Gaussian kernel and using SBM adjacency matrix.



Question 6.3. *How does spectral clustering compare to k-means on the concentric spheres?*

Spectral clustering is better than k-means on SBM. It successfully differentiate the outer sphere (blue) from the inter sphere (red) while k-means can't do that.

Question 6.4. *How does the kernel width affect the output of the clustering?*

I don't see significant difference between spectral clustering with different Gaussian kernel width, the results of clustering are all good, all the 8 clusters are clearly differentiated.

Question 6.5. *Compare the cluster labels of spectral clustering to the filtered noise on the concentric spheres. What is the frequency content of the label vector? Discuss the connection between Spectral clustering and the Graph Fourier Transform.*

With filtered noise, the frequency of ground truth labels becomes low. In spectral clustering, we choose the smallest eigenvalues corresponding to the low frequency eigenvectors. So when we do clustering, we are using the low frequency eigenbasis. Using Graph Fourier Transform to filter the noise has same effect as spectral clustering. We project noise on eigenvector and get a few coefficients, then high frequency components are filtered. The filtered data (coefficients multiply eigenvalue) only includes low frequency signals.

Question 6.6. *In the case that you don't have some pathological geometry (i.e. the concentric spheres), when would you use spectral clustering?*

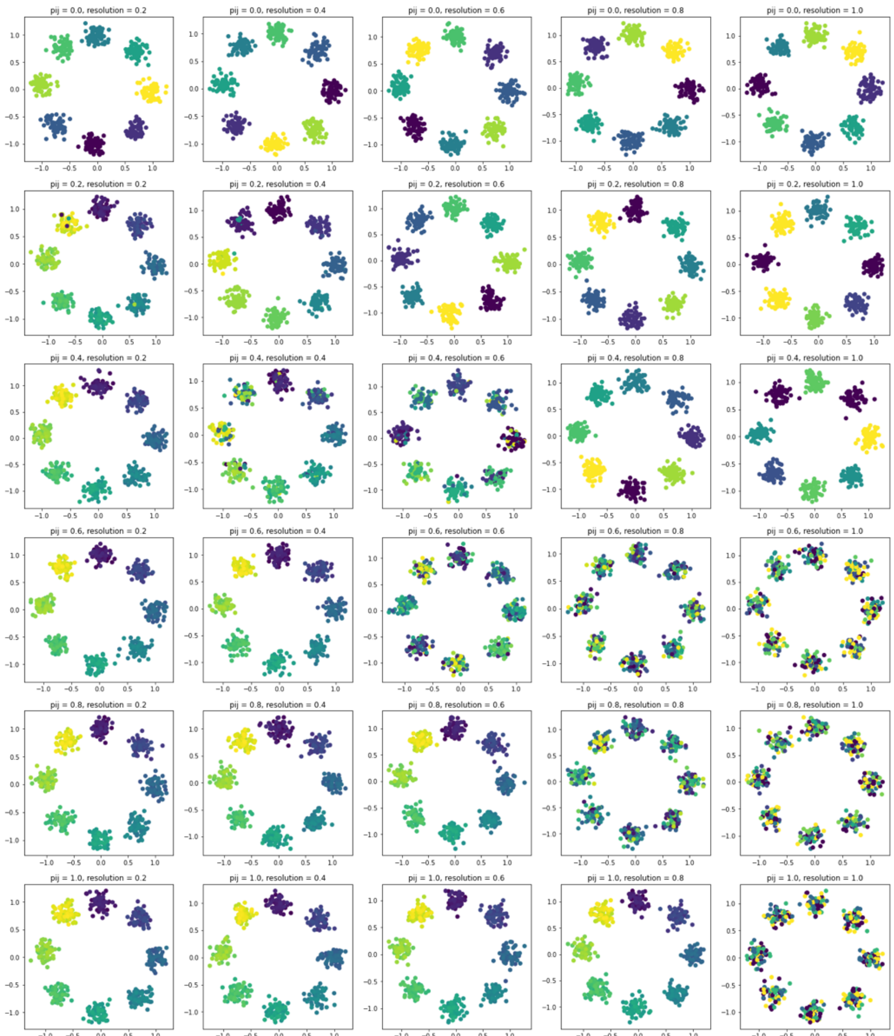
We can use spectral clustering on adjacency matrix or high dimensional data, because we only needs adjacency matrix to do spectral clustering and it performs dimensionality reduction.

7 Using Louvain, and trying out PHATE and tSNE

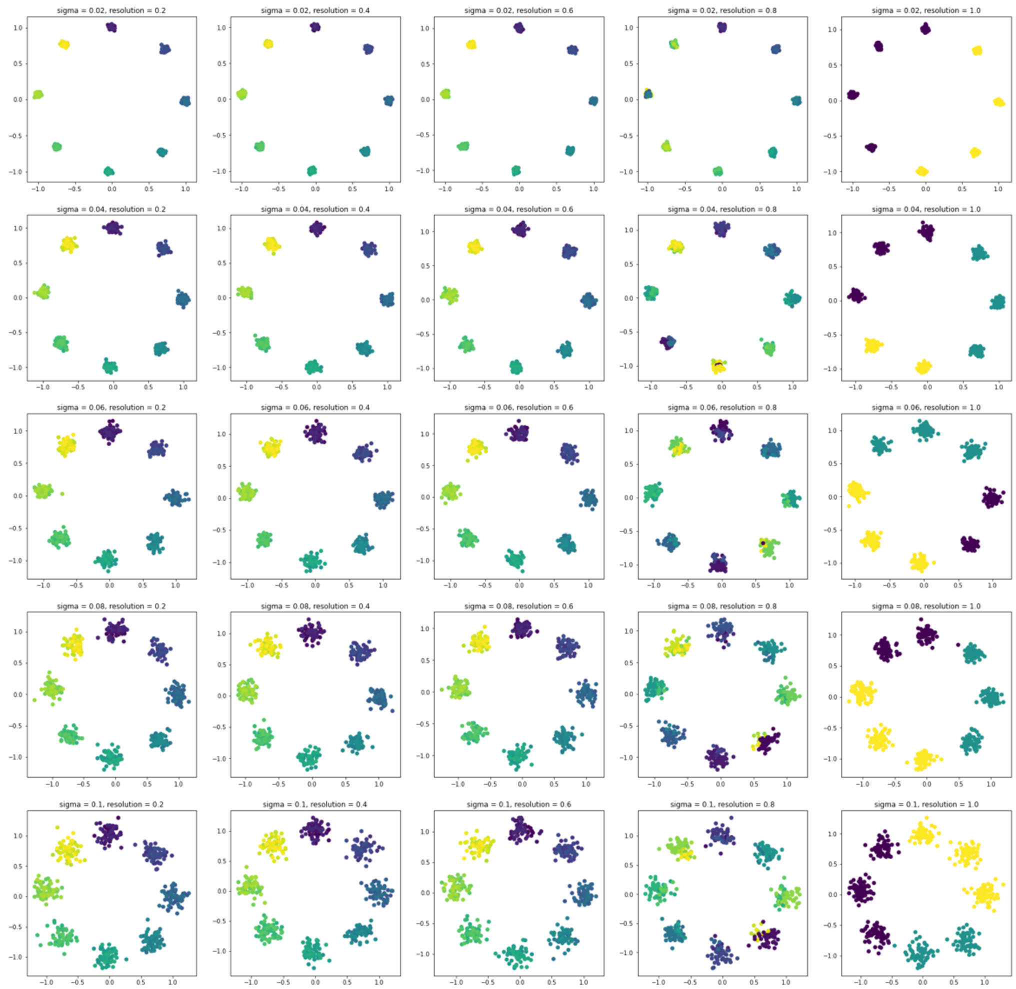
1. First, run Louvain on each of our past examples. Try it on the SBM graphs for a few values of pij. Experiment with Louvain on the GMM, with different sigmas. Lastly, try Louvain on the concentric spheres.

2.Visualize your Louvain clusterings by coloring scatterplots of each model with Louvain's cluster labels, as previously. Try varying Louvain's resolution parameter, while observing the effects on your colored scatterplots.

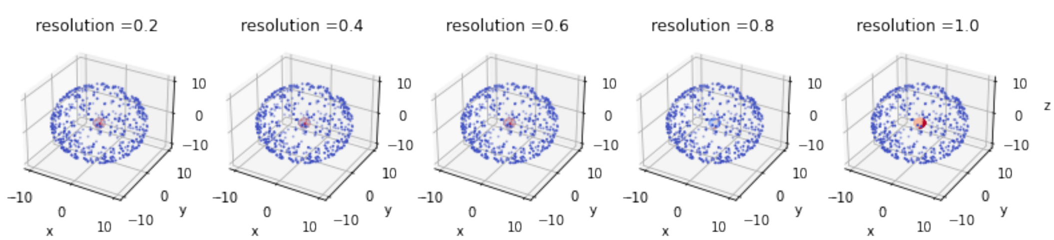
Louvain clustering on SBM (sigma = 0.1)



Louvain clustering on GMM (pij = 0.4)

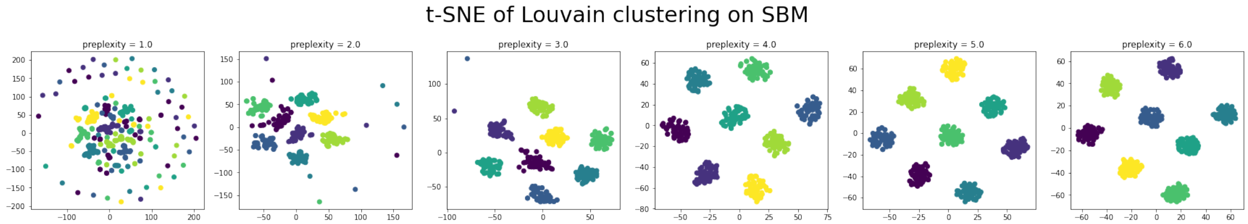


Louvain clustering on two concentric spheres

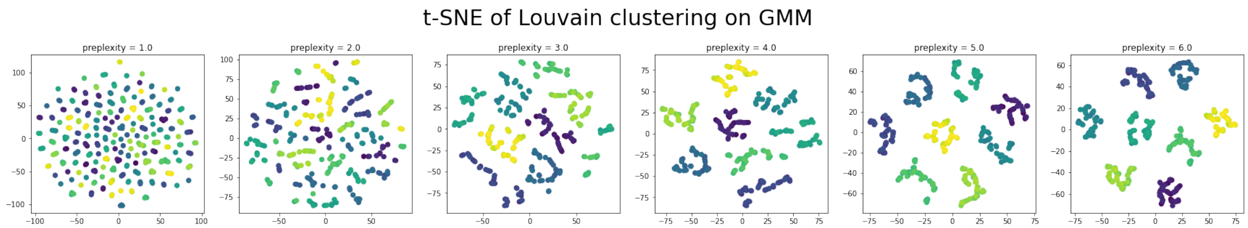


3.Using tSNE, visualize the best clustering you found with Louvain. Create at least six visualizations by varying the parameters of tSNE to find the optimal representation of Louvain's clusters.

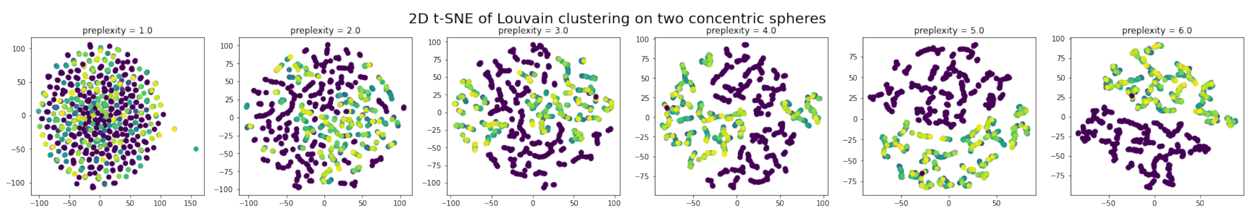
For SBM dataset, the best hyperparameters I found is `pii = 0.8, pij = 0 , sigma = 0.1, resolution = 0.2, perplexity = 6.0`



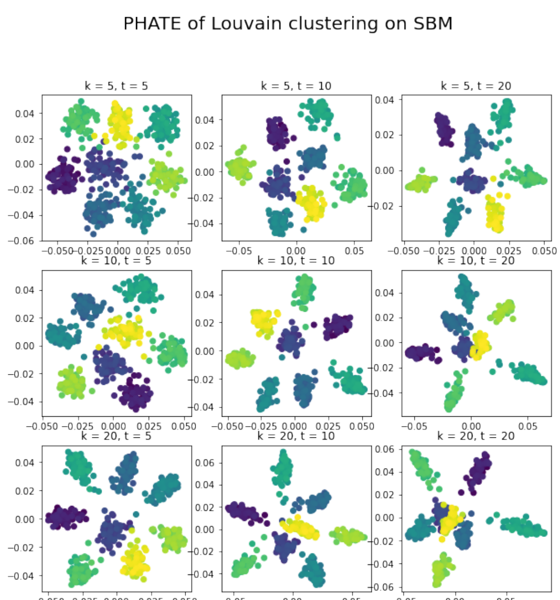
For GMM dataset, the best hyperparameters I found is `pii = 0.8, pij = 0.4 , sigma = 0.1, resolution = 0.2, perplexity = 6.0`

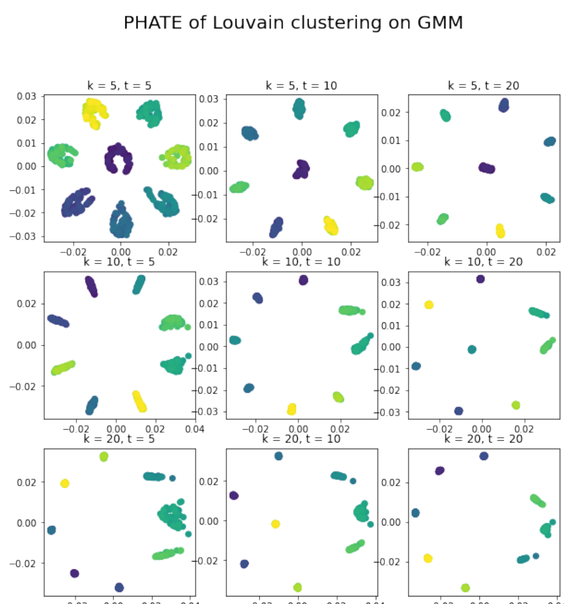


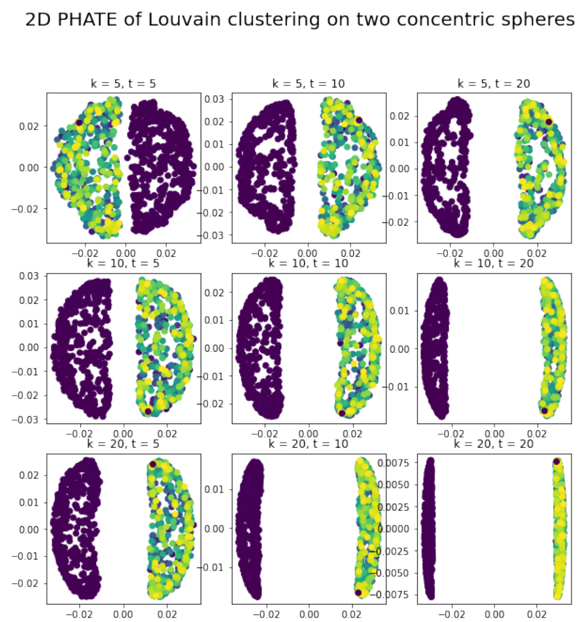
For concentric spheres, the best hyperparameters I found is `resolution = 1.0, perplexity = 6.0`



4.Now, visualize the Louvain clusters using PHATE







Question 7.1. *How does Louvain compare to the output of the two previous clustering methods we tried (spectral and regular k-means)?*

The output of Louvain clustering is similar to spectral clustering, both of them can distinguish 8 different clusters with appropriate combination of hyperparameters (sigma, pij, and resolution) while k-means clustering can't.

Question 7.2. *What is the goal of Louvain?*

The goal of Louvain is to maximize modularity by merging communities in a large graph into smaller communities to make a smaller graph.

Question 7.3. *How does Louvain perform on SBM at various p_{ij} and p_{ii}?*

When pij is small and pii is large, Louvain performs better on SBM.

Question 7.4. *What effects did you observe when varying the parameters of tSNE? Why do you think these changes occurred? Did you find an optimal set of parameters for visualization, or did this depend on the specific clusters?*

The spread and the shape of clusters change when I change perplexity of t-SNE.   Because perplexity is related to the number of nearest neighbors that is used in other manifold learning algorithms. Changing perplexity will lead to change of cost function, resulting in the change of pairwise distances between similar data points and between dissimilar data points. I think it depends on specific clusters, because for the 3 datasets SBM, GMM and concentric spheres with different hyperparameters, the optimal perplexity is different.

Question 7.5. *What effects did you observe when varying the parameters of PHATE? Why do you think these changes occurred? Was there an optimal set of parameters for all visualizations?*

The size of cluster changes when I change knn and t of PHATE. Because knn is the number of nearest neighbors on which to build kernel. t is the time step of random walk, the power to which the diffusion operator is powered. This sets the level of diffusion. I think the optimal set of params for visualization depends on specific clusters, because for the 3 datasets SBM, GMM and concentric spheres with different hyperparameters, the optimal perplexity is different.

Question 7.6. *What differences did you notice between tSNE and PHATE? Theoretically, what do you think accounts for them?*

PHATE is more robust to capture both global and local structure of data while t-SNE only focuses on local structure of data at the cost of shattering the global picture. This is particularly significant when we compare visualization results on concentric spheres dataset. t-SNE can't distinguish 2 spheres clearly while PHATE creates a clear boundary between 2 spheres with different colors (purple vs. yellow). Because t-SNE is aimed to minimize Kullback–Leibler divergence, which push near points further near and faraway points further faraway, while PHATE use kernel to preserve local structure and use affinity to preserve global structure by "diffuse" through the data via a Markov random-walk.

Question 7.7. *Why is the Gaussian mixture model a good data set to use to test the clustering algorithm?*

Because we know the probability for which each point belongs to a cluster in GMM. If the probability calculated by a clustering *algorithm* is similar as ground truth, then it is a decent *algorithm*.

Question 7.8. *What modifications to the mixture model would you make to further test the clustering algorithms?*

I may filter the clusters with low probabilities and only keep clusters with high probabilities (e.g., 90%) after applying Gaussian kernel.

Question 7.9. *What are other synthetic data sets you would use to test clustering algorithms?*

A chain of interconnected rings which are orthogonal to each other like the figure below.

8 Retinal Bipolar dataset

1.Load the Retinal Bipolar dataset and its metadata. (Note that the metadata actually contains 26 cluster labels, though Shekhar only found 15 of them to be biologically significant.)

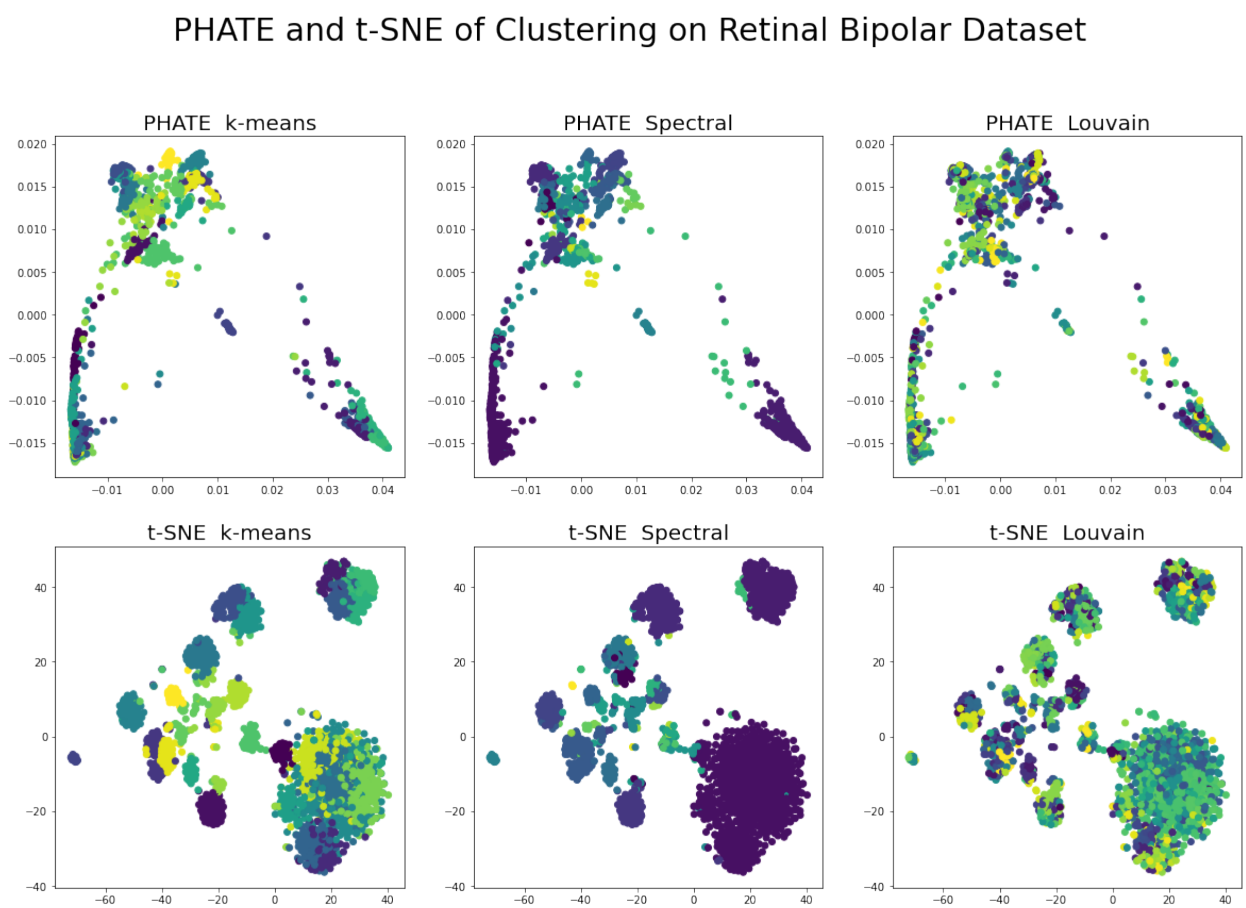
2.The unaltered dataset has over 15,000 columns corresponding to 15,000 markers for each of the 21,000 cells. To make the computations feasible on your laptop, we'll apply PCA to the columns of the dataset, and project each cell onto just the first 100 PCA components.

3.21,000 cells is still quite a few, so we'll subsample down to 3000 cells.

4.Build a graph from the dimensionally-reduced data using an adaptive Gaussian kernel with k = 10.

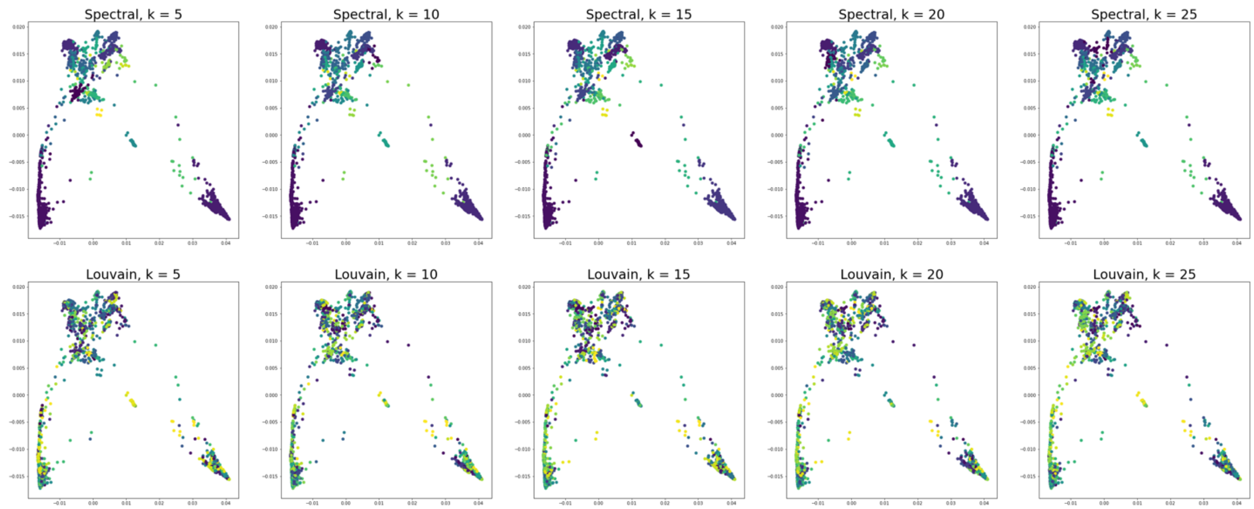
5.Cluster the data using k-means, Spectral Clustering and Louvain.

6.Visualize the dimensionally-reduced data using PHATE, coloring the points according to each of the above clusterings. Produce similar visualizations using tSNE. What differences do you notice between the visualization techniques with this dataset?

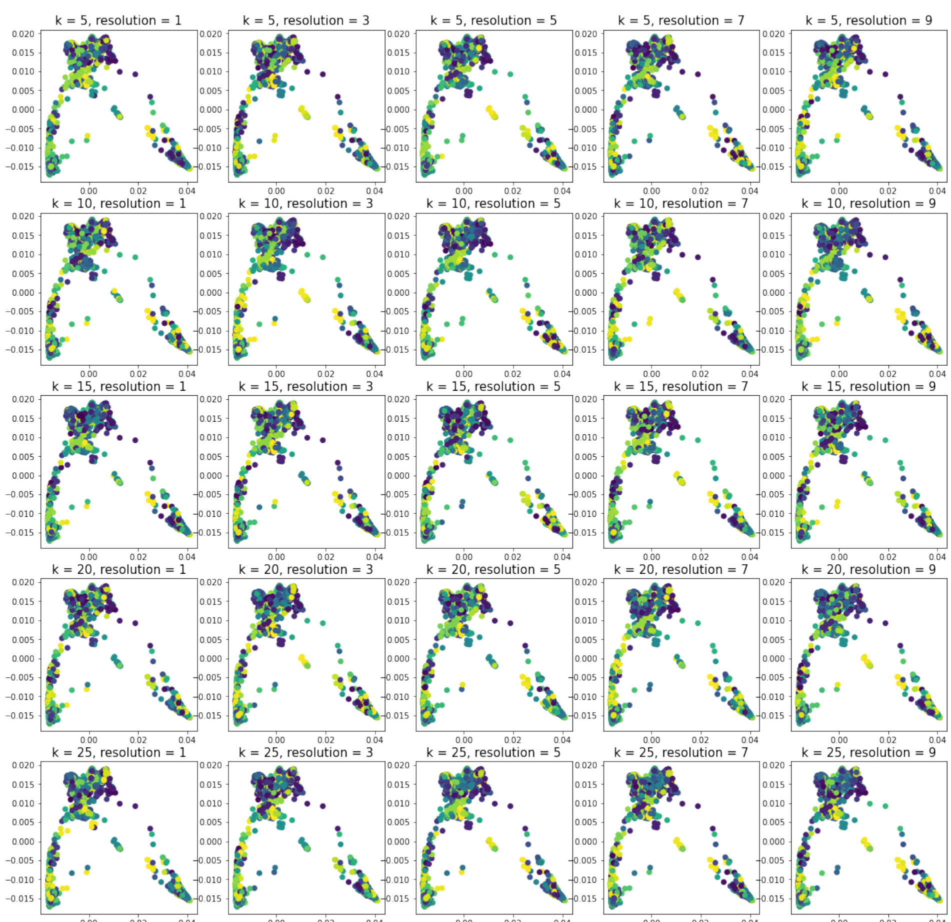


7.Try varying the kernel parameter k to obtain different cluster assignments. Additionally vary the Louvain parameters. Visualize each of these new clusterings with a recolored PHATE plot.

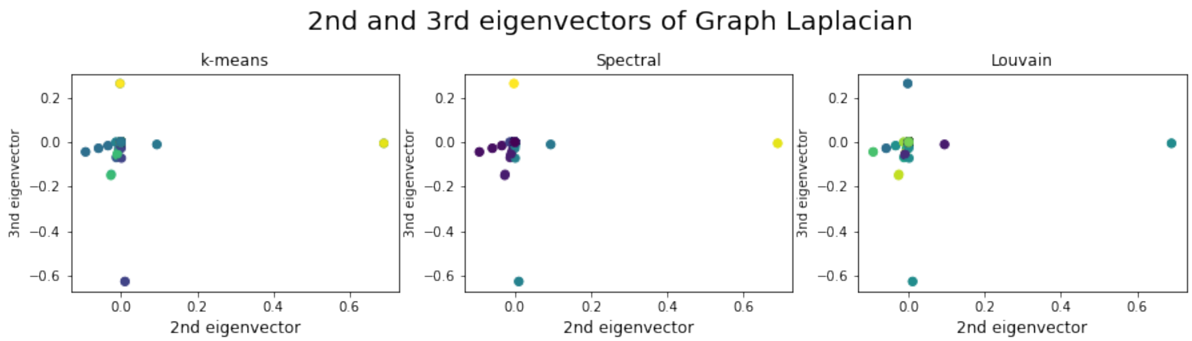
PHATE of Clustering on Retinal Bipolar Dataset



PHATE of Louvain Clustering on Retinal Bipolar Dataset

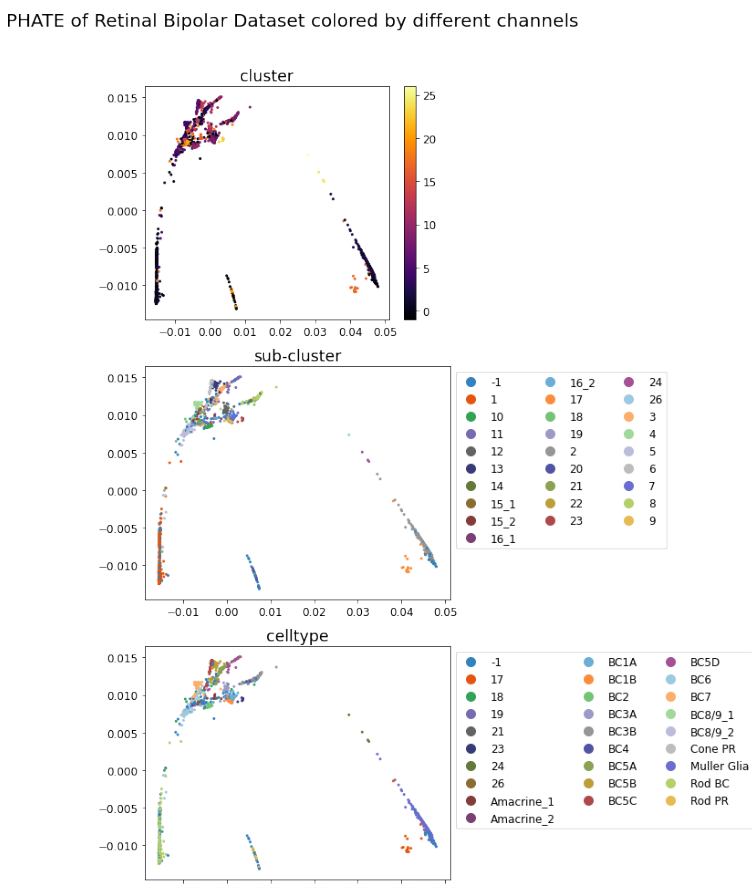


8.Plot the second and third eigenvectors of the Graph Laplacian. Color the plot by your cluster assignments.



9.Try coloring the PHATE plot using different channels in the data (e.g. different columns of the metadata matrix). Can you find any channels that seem to correspond to your cluster assignments?

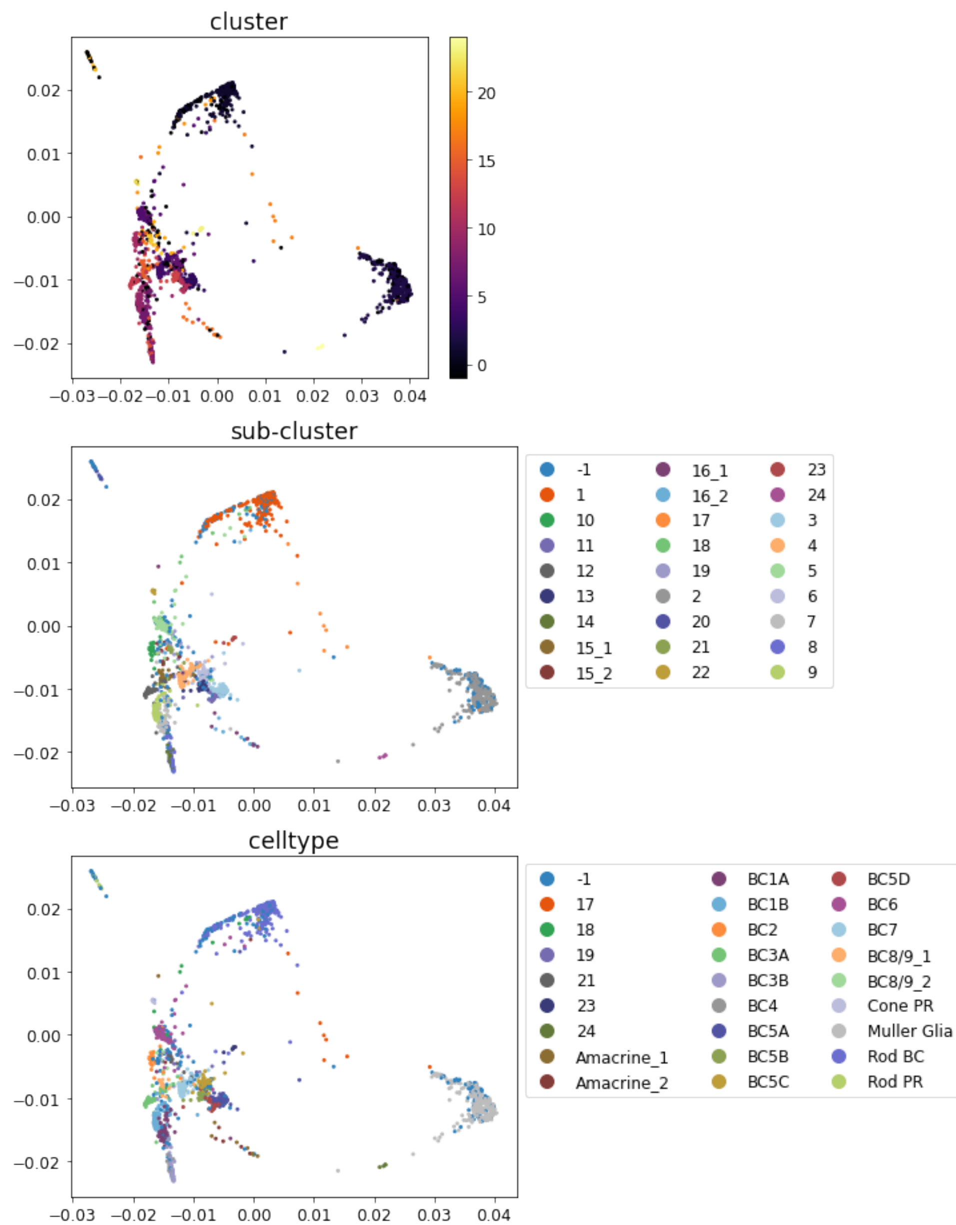
I found that channels Rod BC, Muller Glia, BC5A, BC5C, BC5B, BC6, BC1A, BC2, BC3A seem to correspond to my cluster assignments.



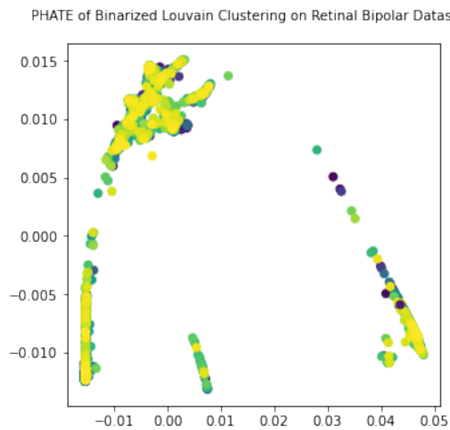
10.Apply an ideal low-pass filter to some of the channels you found above, treating the gene expression as a signal over the graph. Rerun your visualizations using these new, filtered channels to color the PHATE plot. How much has changed? Do any of the denoised channels better represent the true clustering?

The signals on the denoised channels are magnified. Among them, Rod BC, BC5A, Muller Glia, BC6, BC1A, BC5C better represent the true clustering

PHATE of Retinal Bipolar Dataset colored by channels via low pass filter



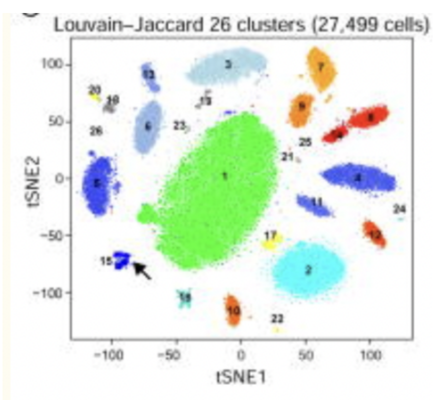
11.Binarize the kernel such that it is no longer a weighted graph (i.e., if a value is greater than some threshold, it is 1, otherwise, it is 0). Try rerunning Louvain with this binarized kernel.

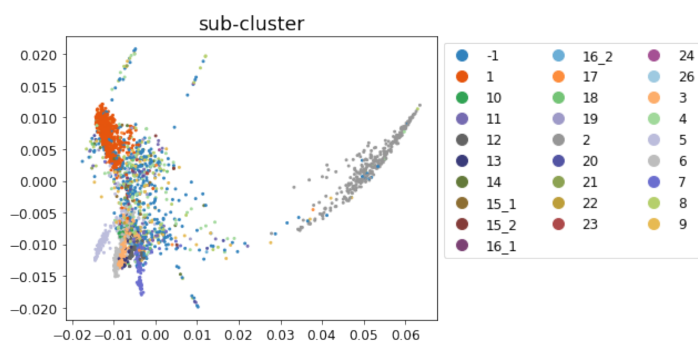


12.Finally, compare your clusters to those of Shekhar et al. by coloring your PHATE visualizations with the cluster number provided in the metadata file. Describe any relationships you notice between these clusters and the clusters you've obtained with k-means, Louvain, and Spectral Clustering.

In my PHATE visualization, points with sub-cluster id 1, 2, 3, 4, 6, 7 forms clusters with decent size respectively, which is consistent with Shekhar's paper. Spectral clustering performs better than both k-means and Louvain, I can see at least 6 clusters with decent size.

The left figure below is t-SNE visualizations of clusters in the Shekhar's paper, the right figure below is my PHATE visualization of clusters.





Question 8.1. *How many clusters did Louvain produce? How many clusters did PHATE and tSNE suggest? To what extent did different clustering methods parameters affect this number?*

I can't see clearly how many clusters Louvain produced. t-SNE suggests 18 clusters. I can't see clearly how many clusters PHATE suggests. The different parameters don't affect too much on the number of clusters.

Question 8.2. *In a biological setting, how would you interpret the different clusters? More specifically, howwould you interpret any variation you noticed from the 15 clusters described by Shekhar et al?*

Different clusters are gene variation in single cell level.

Question 8.3. *What did you notice when you plotted the eigenvectors of the Graph Laplacian?*

They are orthogonal to each other

Question 8.4. *What is the effect of low-pass filtering a feature on the graph? In what context could this be useful for data analysis?*

The low-pass filter decays the low frequency signals which are often noises on the graph while magnify the high frequency signals on the graph which tend to be real signals. When noises are low-frequency signals or we want to smooth the signals.

Question 8.5. *How does binarization of the kernel affect your clustering? Why might this be the case?*

Binarization of kernel reduces the number of clusters because it forces the entries in the affinity matrix to be either 0 or 1 to simplify the connectivity between clusters.