ml_project

March 2, 2020

```
In [38]: import pandas as pd
         import numpy as np
         import scipy.linalg as la
         import scipy.stats as stats
         import statsmodels.api as sm
         from sklearn import linear_model, model_selection, metrics
         import sklearn
         import plotly.graph_objs as go
         import matplotlib.pyplot as plt
         from sklearn.datasets import load_breast_cancer, load_digits, load_diabetes
         from sklearn.model_selection import train_test_split
         from sklearn.discriminant_analysis import LinearDiscriminantAnalysis as LDA
         from sklearn.neighbors import KNeighborsClassifier as KNC
         from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis as GDA
         from sklearn.metrics import roc_auc_score as RAS
         from sklearn.metrics import roc_curve as ROC
         import seaborn as sns
         from sklearn.manifold import TSNE
         import umap
```

```
   jupyter nbconvert --to latex "Race and Incarceration in America.ipynb"
--TagRemovePreprocessor.remove_input_tags='{"hide_input"}'; pdflatex "Race and
Incarceration in America"
```

1 Introduction

Last semester I took a look a data set containing the information of more than 7.5 million individuals that have been processed by the criminal justice system. I found that racial minorities were more likely to receive extreme sentences, agreeing with existing research around bias in the criminal justice system. In this project I will be exploring the data from a machine learning perspective. My goal is to determine if this data can be classified in such a way that is predictive of race.

2 Possible questions

can we use ml techniques to correctly classify this data? which ones fail and why? can we create a predictive model for sentence lengths? should sentencing be offloaded to a ml algorithm.

```
In [14]: chunksize = 100000
         rdf = pd.read_csv('regression_df.csv', chunksize=chunksize)
```

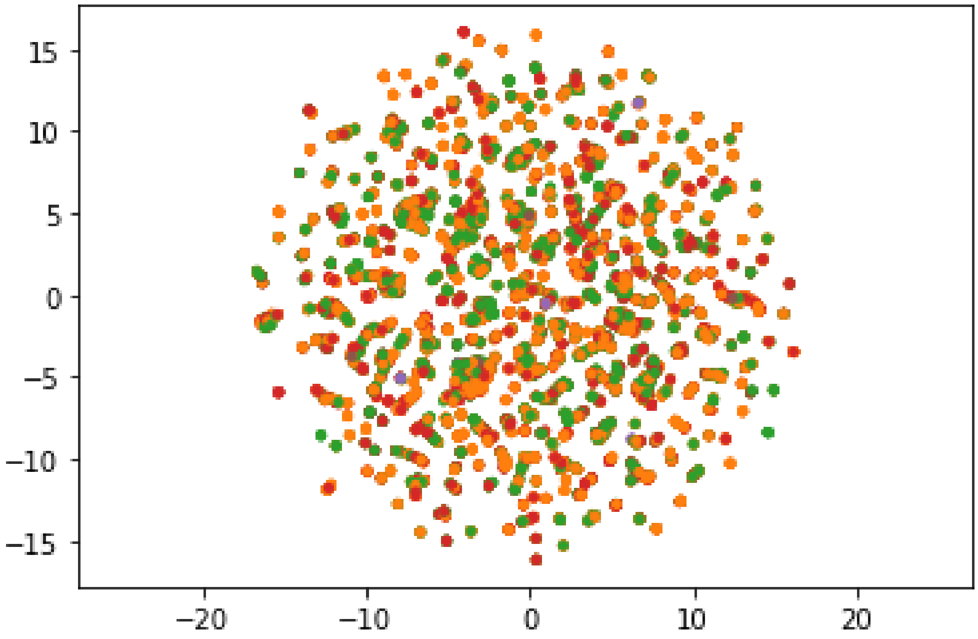
```
In [39]: # reducer = umap.UMAP()
         # i = 0
         # for item in rdf:
         #     if i > 3:
         #         break
         #     else:
         #         i += 1

         #     data = item.drop(['Unnamed: 0','IDENTIFIER','LATEST ADMISSION DATE'], axis=1)
         #     target = data[['AMER IND','ASIAN','BLACK','HISPANIC']]
         #     data = data.drop(['AMER IND','ASIAN','BLACK','HISPANIC'], axis=1)
         #     trans = reducer.fit_transform(data)

         # #     create maskings
         #     black = target.BLACK == 1
         #     asian = target.ASIAN == 1
         #     hisp = target.HISPANIC == 1
         #     ind = target['AMER IND'] == 1
         #     white = ~(black + asian + hisp)

         #     plt.scatter(trans[:,0][black],trans[:,1][black],label='black',marker='.')
         # #     plt.legend()
         # #     plt.show()
         #     plt.scatter(trans[:,0][white],trans[:,1][white],label='white',marker='.')
         # #     plt.legend()
         # #     plt.show()
         #     plt.scatter(trans[:,0][hisp],trans[:,1][hisp],label='hisp',marker='.')
         # #     plt.legend()
         # #     plt.show()
         #     plt.scatter(trans[:,0][ind],trans[:,1][ind],label='ind',marker='.')
         # #     plt.legend()
         # #     plt.show()
         #     plt.scatter(trans[:,0][asian],trans[:,1][asian],label='asian',marker='.')
         #     plt.legend()
         #     plt.show()
```

```
In [29]: cols = ['RACE','GENDER','AGE','OFFENSE','FACILITY','DETAINER','SENTENCE DAYS']
         df = pd.read_csv('individuals.csv',usecols=cols)
```

```
In [45]: samp = df.sample(20000)
         samp.RACE = pd.factorize(samp['RACE'])[0] + 1
         samp.GENDER = pd.factorize(samp['GENDER'])[0] + 1
         samp.OFFENSE = pd.factorize(samp['OFFENSE'])[0] + 1
         samp.DETAINER = pd.factorize(samp['DETAINER'])[0] + 1
         samp.FACILITY = pd.factorize(samp['FACILITY'])[0] + 1
```

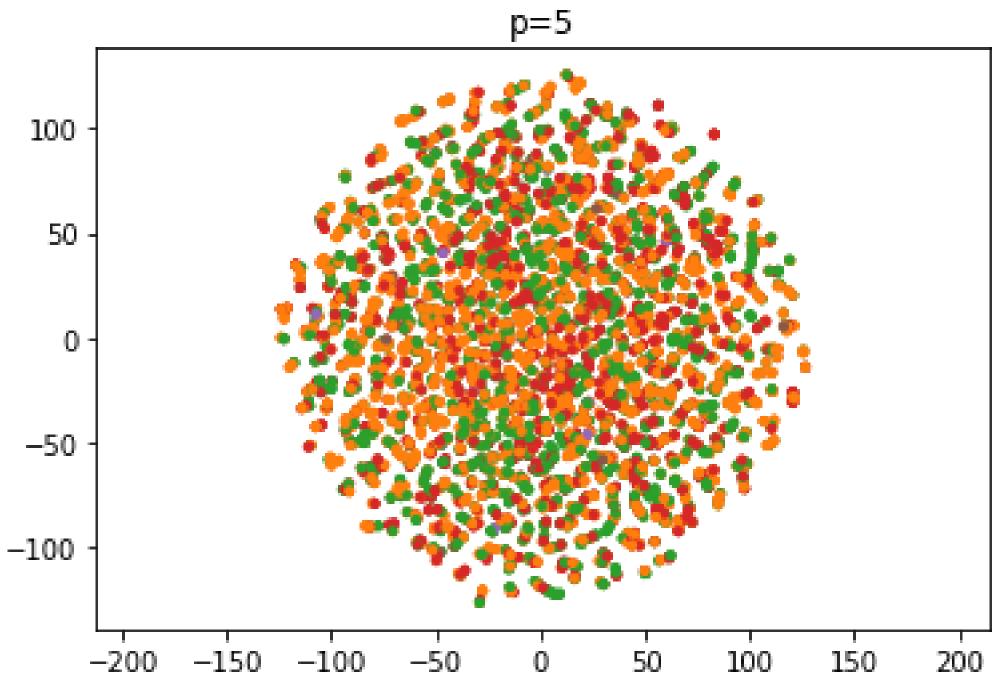
```
samp_y = samp.RACE
samp_X = samp[['GENDER','AGE','OFFENSE','FACILITY','DETAINER','SENTENCE DAYS']]

trans = reducer.fit_transform(samp_X)

plt.scatter(trans[:, 0], trans[:, 1], c=[sns.color_palette()[x] for x in samp_y], marl
plt.gca().set_aspect('equal', 'datalim')
plt.show()
```

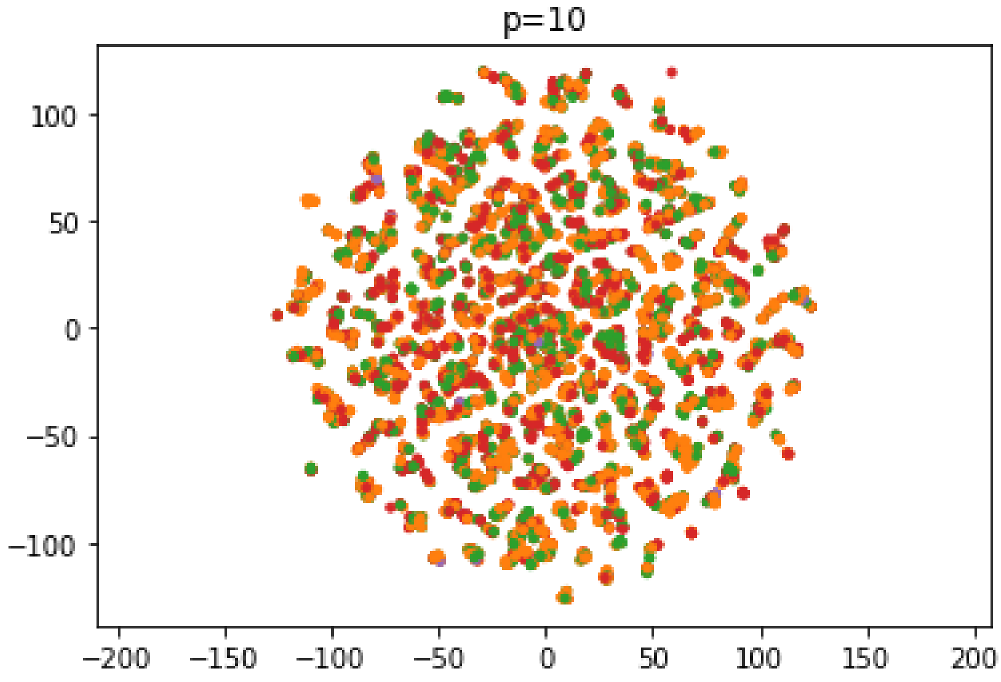
```
/home/ethan/.local/lib/python3.7/site-packages/numba/compiler.py:602: NumbaPerformanceWarning:


The keyword argument 'parallel=True' was specified but no transformation for parallel executior

To find out why, try turning on parallel diagnostics, see http://numba.pydata.org/numba-doc/lat

File "../../../.local/lib/python3.7/site-packages/umap/nndescent.py", line 47:
    @numba.njit(parallel=True)
    def nn_descent(
    ^



/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 50 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```

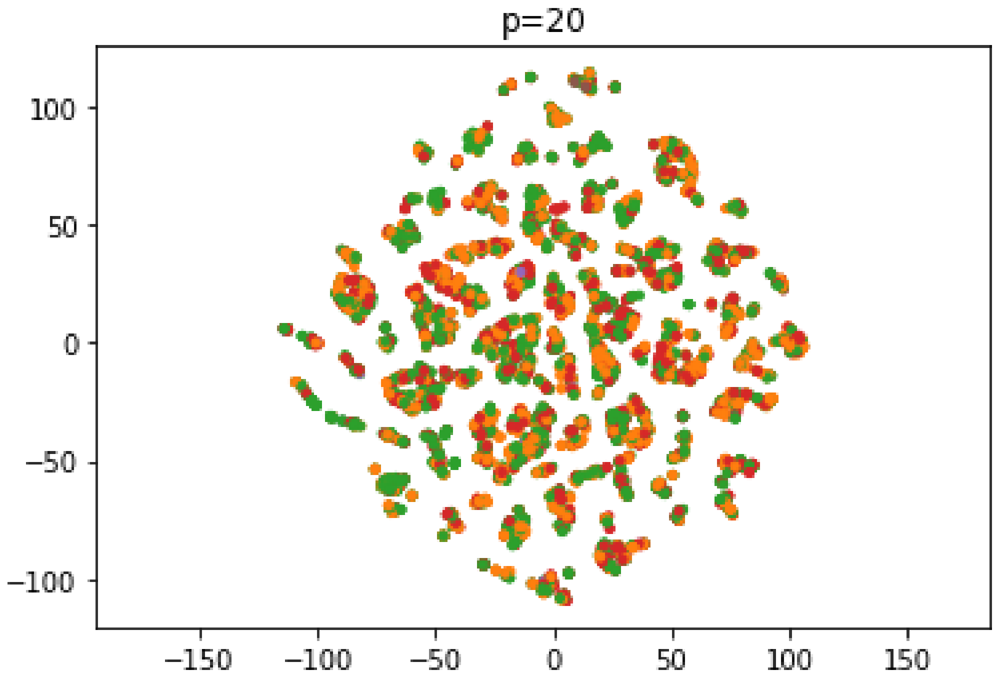


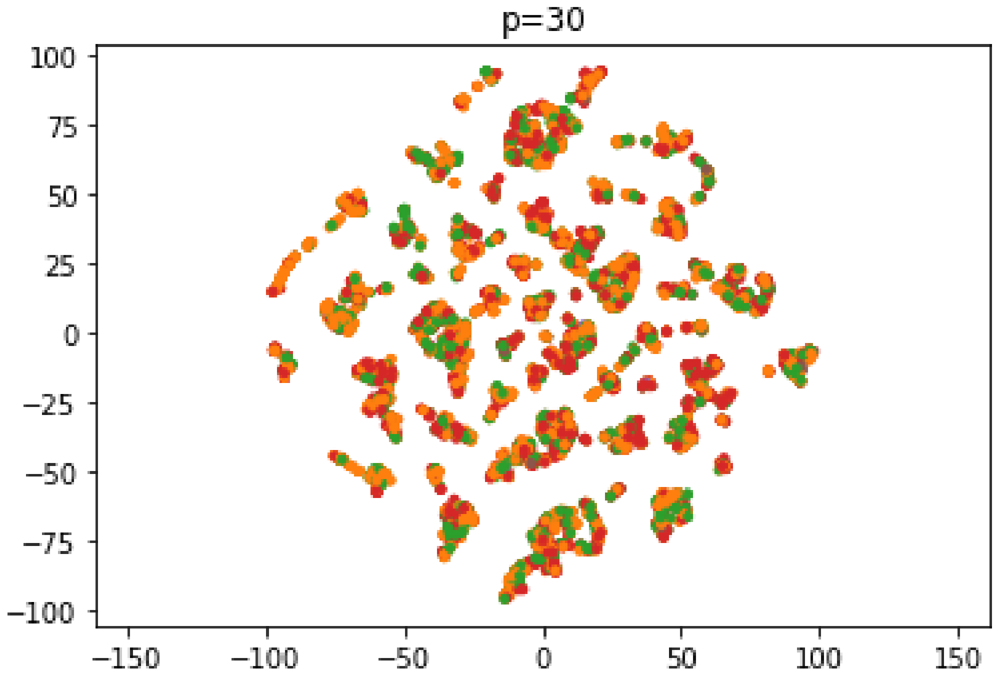
It seems that UMAP doesnt work

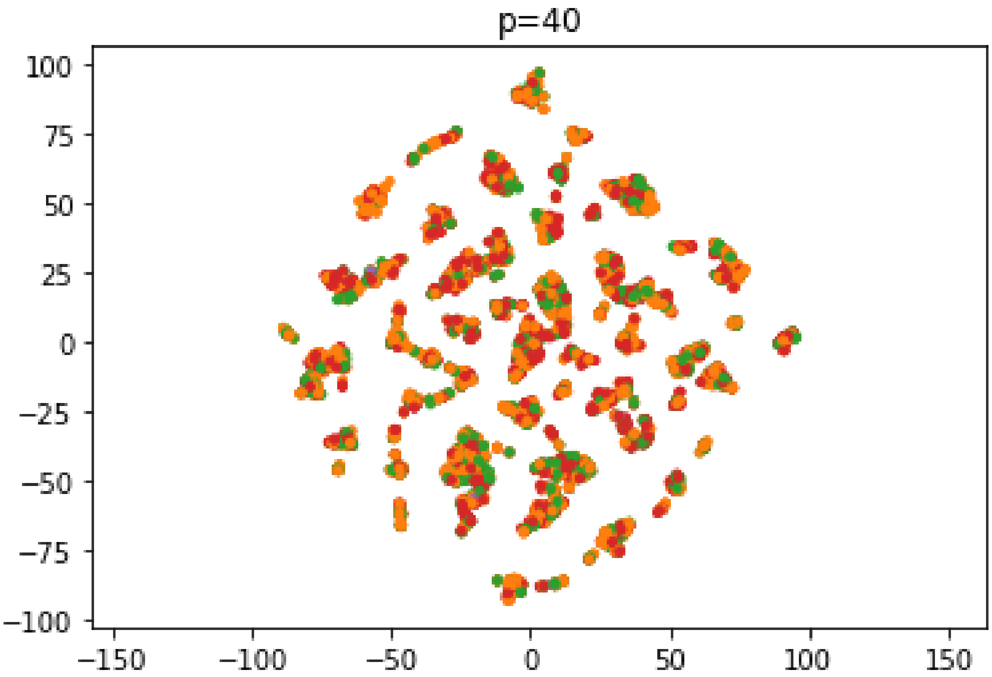
```
In [47]: # params
         ps = [5,10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190,200]

         for p in ps:

             samp = df.sample(8000)
             samp.RACE = pd.factorize(samp['RACE'])[0] + 1
             samp.GENDER = pd.factorize(samp['GENDER'])[0] + 1
             samp.OFFENSE = pd.factorize(samp['OFFENSE'])[0] + 1
             samp.DETAINER = pd.factorize(samp['DETAINER'])[0] + 1
             samp.FACILITY = pd.factorize(samp['FACILITY'])[0] + 1
             samp_y = samp.RACE
             samp_X = samp[['GENDER','AGE','OFFENSE','FACILITY','DETAINER','SENTENCE DAYS']]



             trans = TSNE(perplexity=p).fit_transform(samp_X)

             plt.scatter(trans[:, 0], trans[:, 1], c=[sns.color_palette()[x] for x in samp_y],
             plt.gca().set_aspect('equal', 'datalim')
             plt.title(f'p={p}')
             plt.show()
```

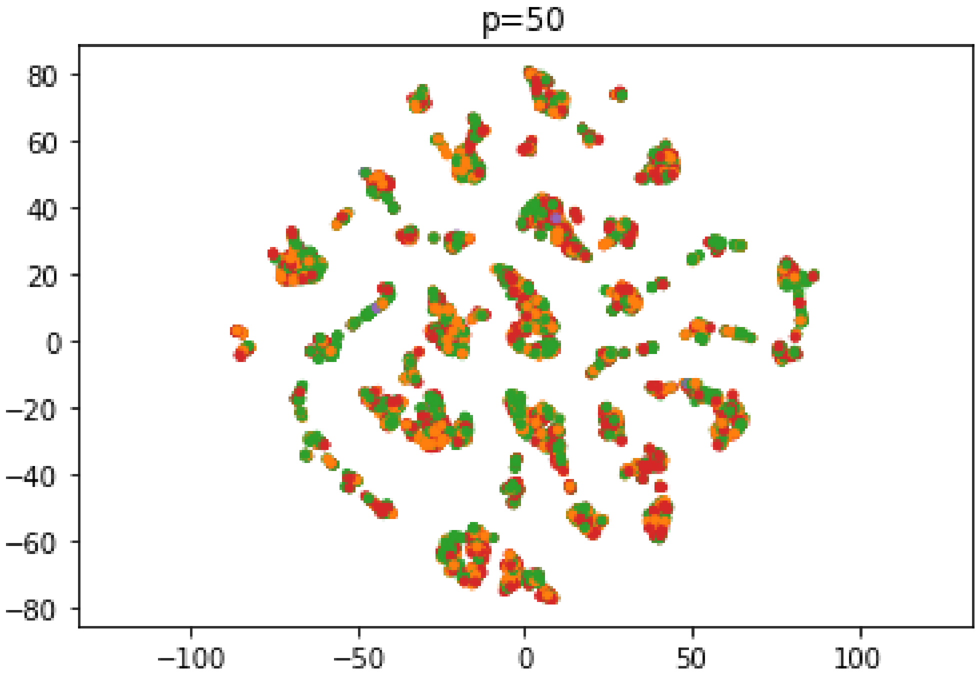

p=5
100
50
0
−50
−100
−200
−150
−100
−50
0
50
100
150
200

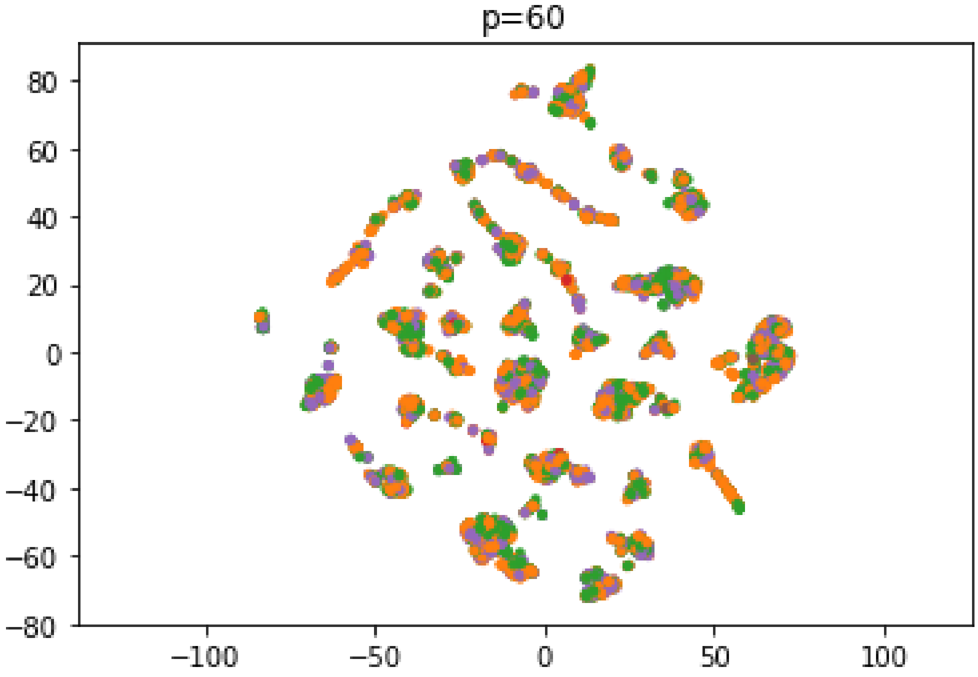

p=10
100
50
0
−50
−100
−200
−150
−100
−50
0
50
100
150
200

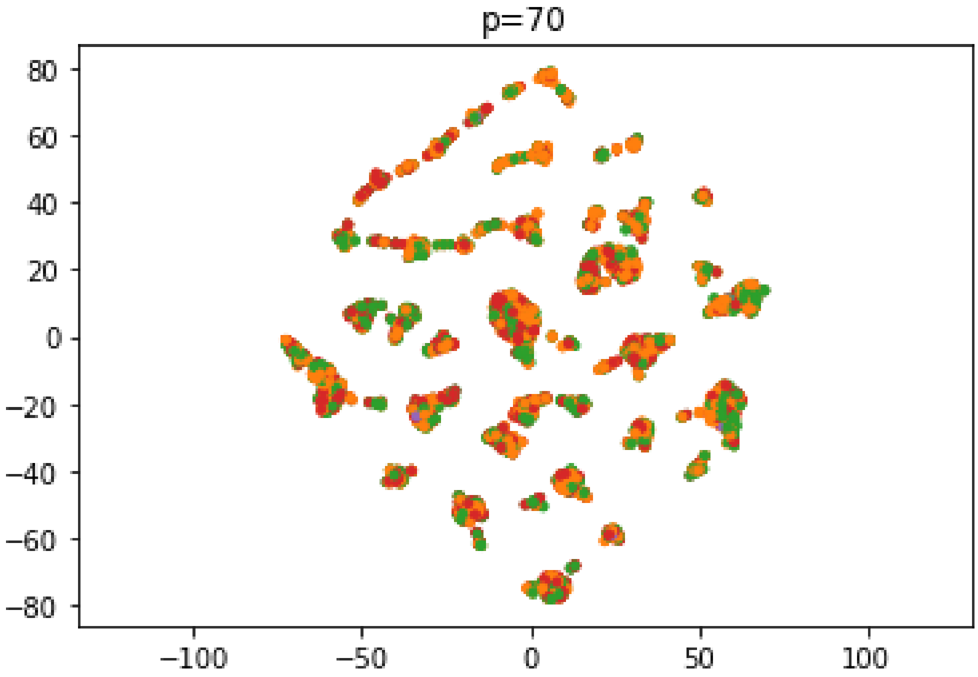

p=20
100
50
0
−50
−100
−150
−100
−50
0
50
100
150

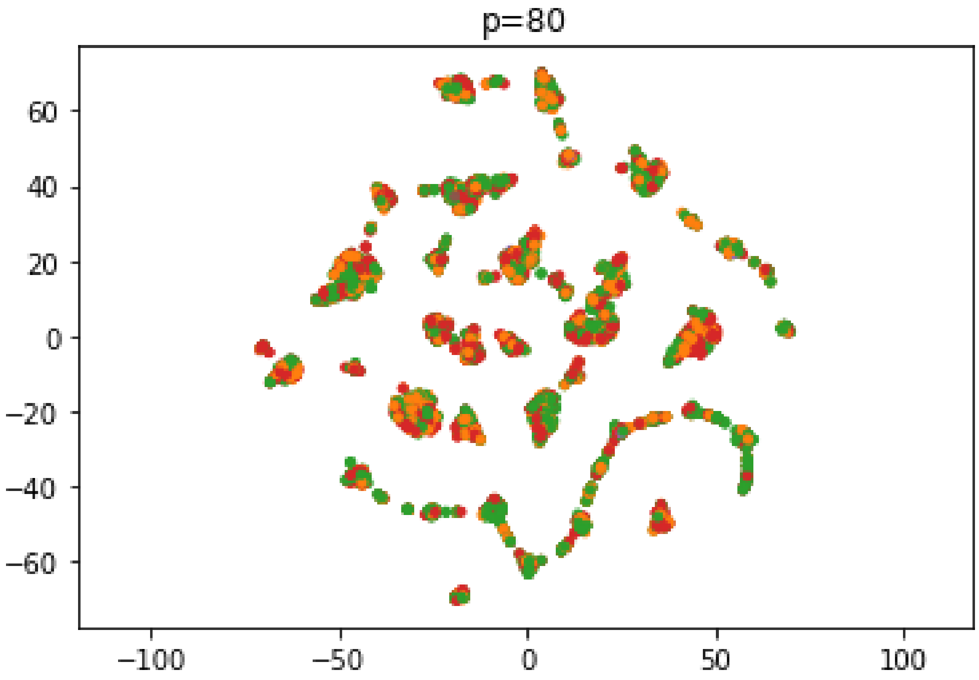

p=30
100
75
50
25
0
−25
−50
−75
−100
−150
−100
−50
0
50
100
150

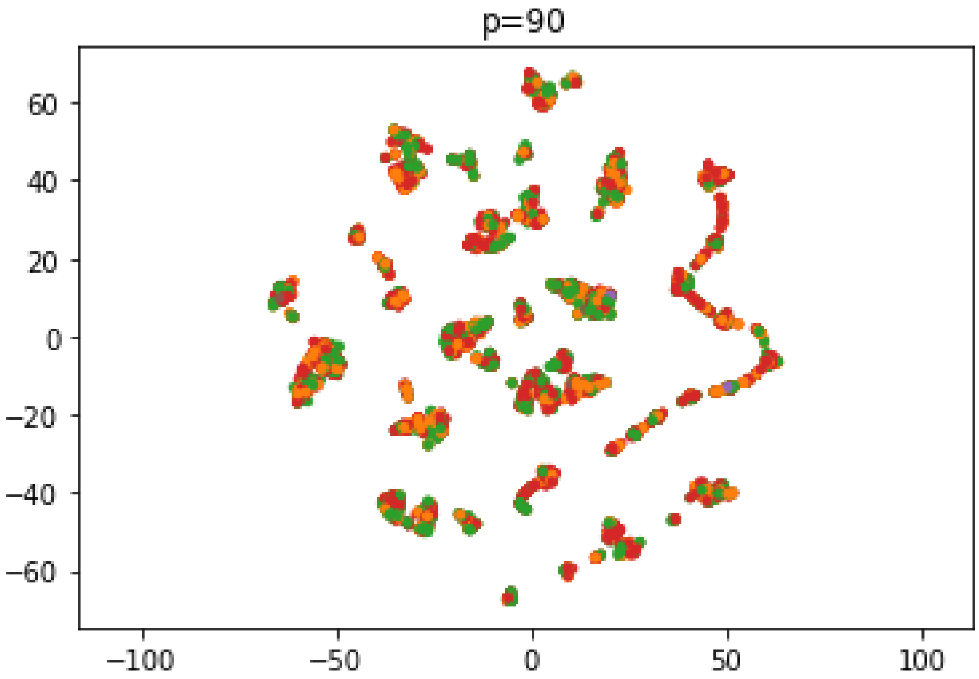

p=40
100
75
50
25
0
−25
−50
−75
−100
−150
−100
−50
0
50
100
150

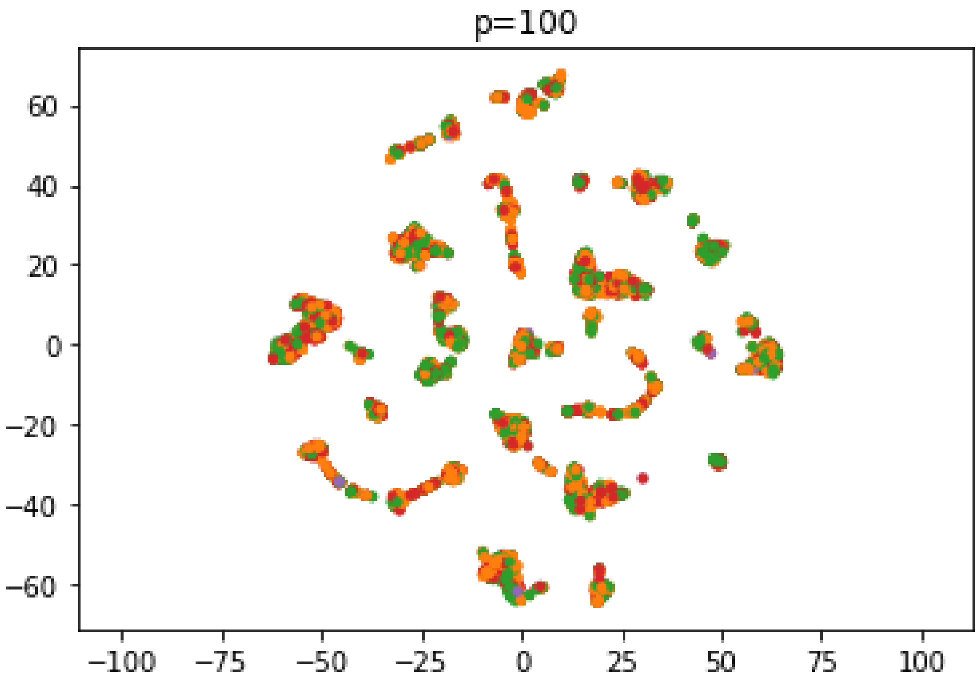

p=50
80
60
40
20
0
−20
−40
−60
−80
−100
−50
0
50
100

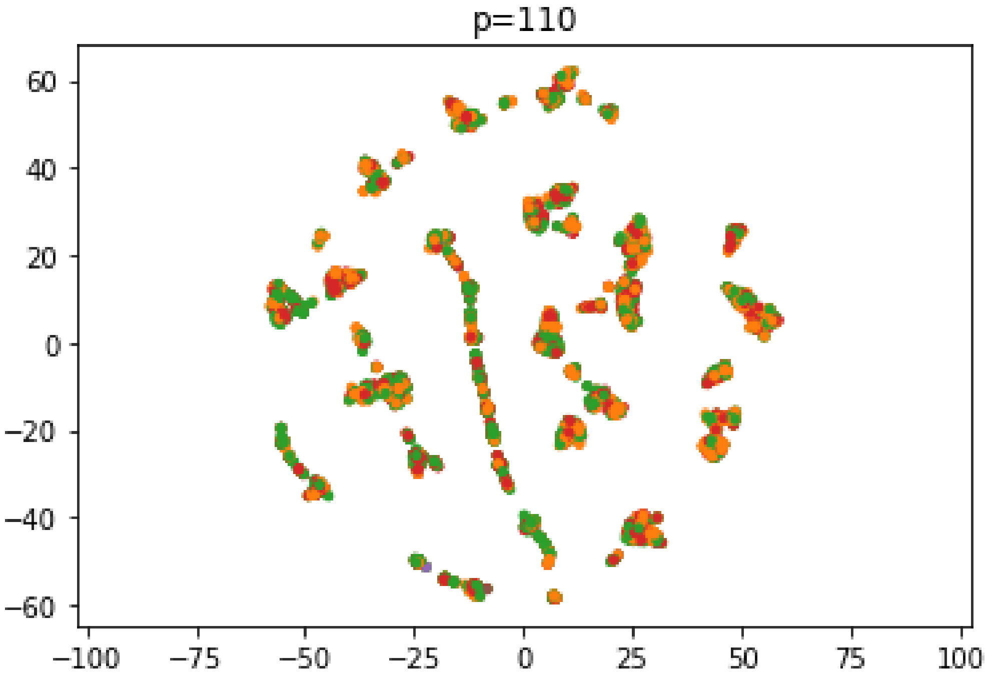

p=60
80
60
40
20
0
−20
−40
−60
−80
−100
−50
0
50
100

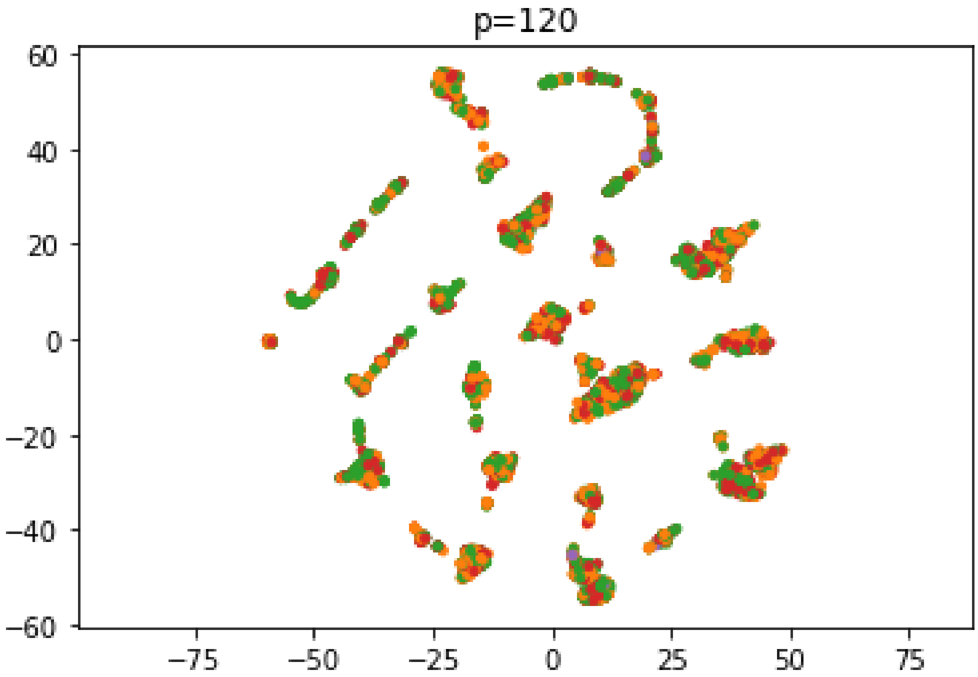

p=70
80
60
40
20
0
−20
−40
−60
−80
−100
−50
0
50
100

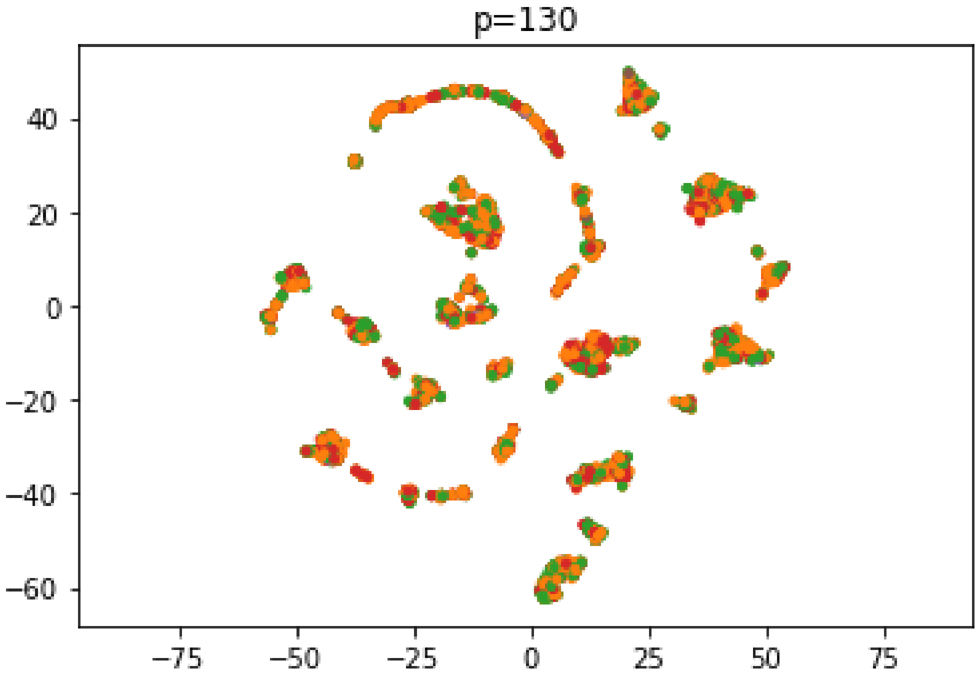

p=80
60
40
20
0
−20
−40
−60
−100
−50
0
50
100

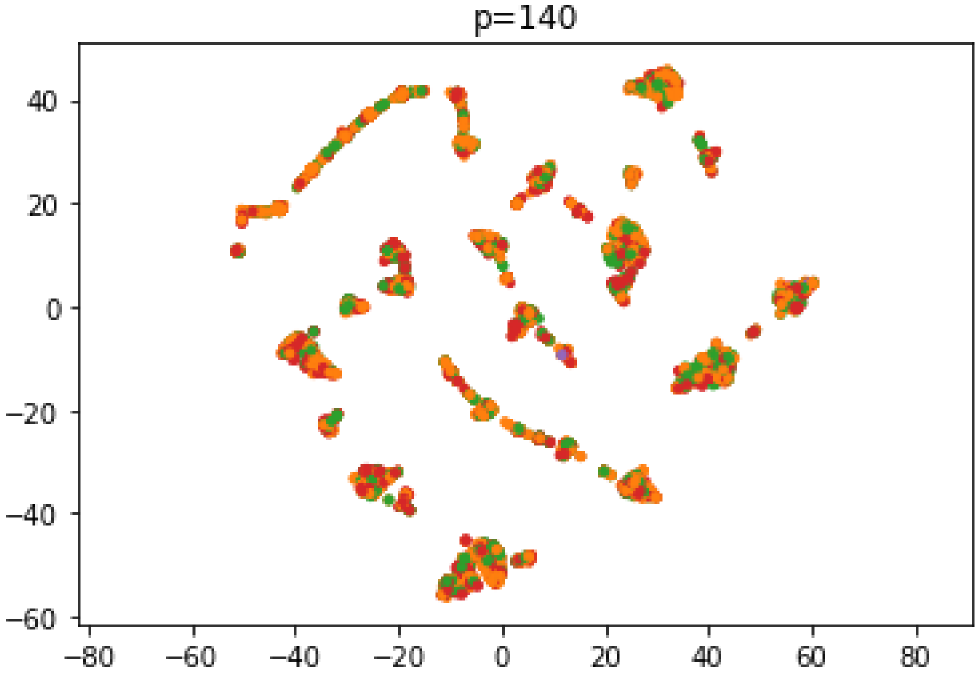

p=90
60
40
20
0
−20
−40
−60
−100
−50
0
50
100

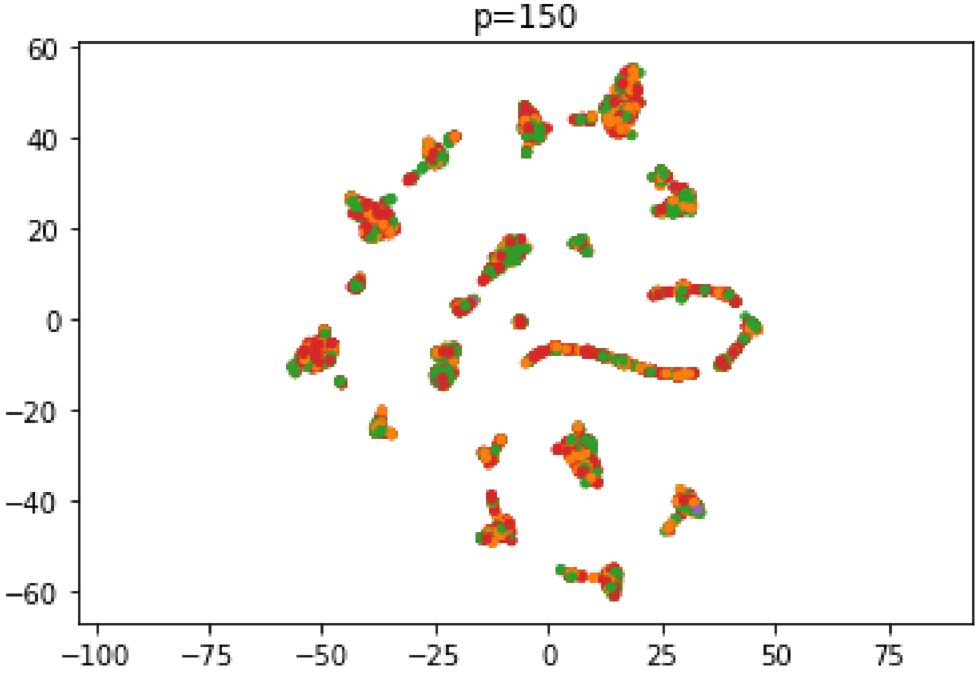

p=100
60
40
20
0
−20
−40
−60
−100
−75
−50
−25
0
25
50
75
100

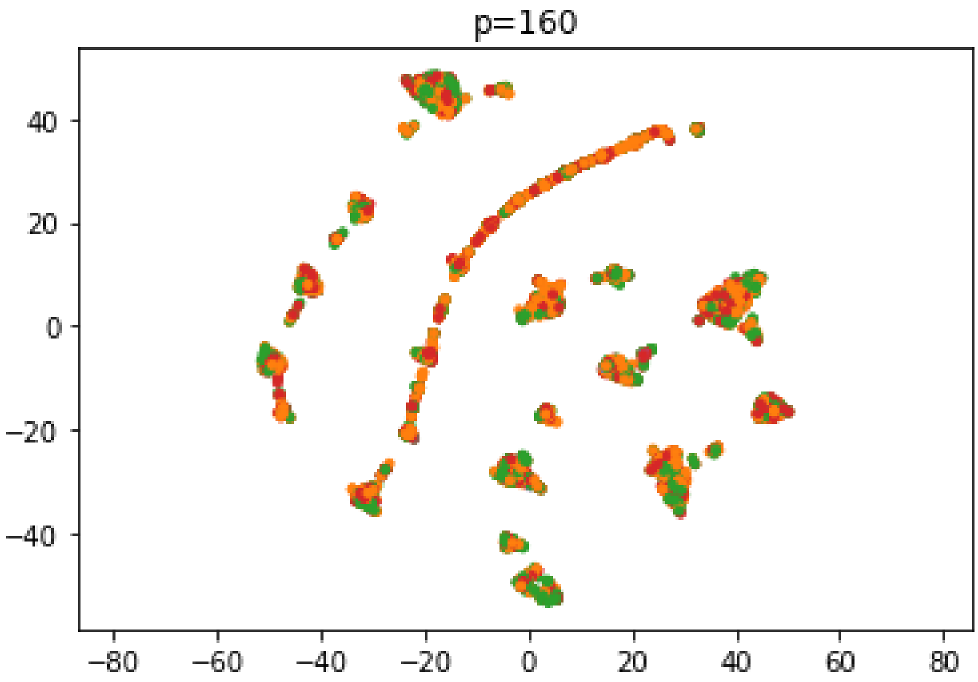

p=110
60
40
20
0
−20
−40
−60
−100
−75
−50
−25
0
25
50
75
100

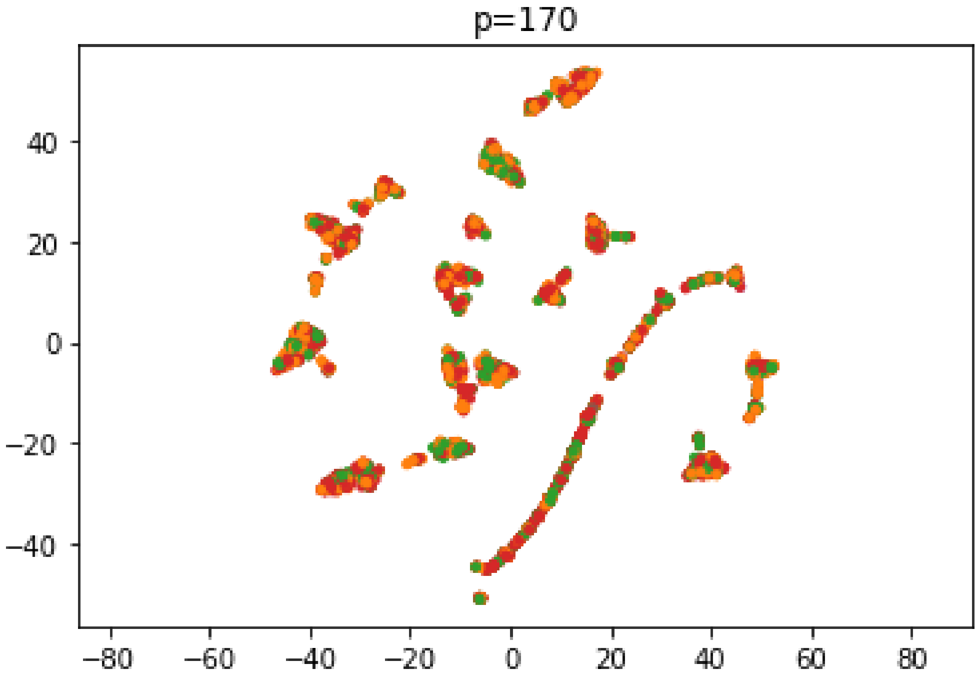

p=120
60
40
20
0
−20
−40
−60
−75
−50
−25
0
25
50
75

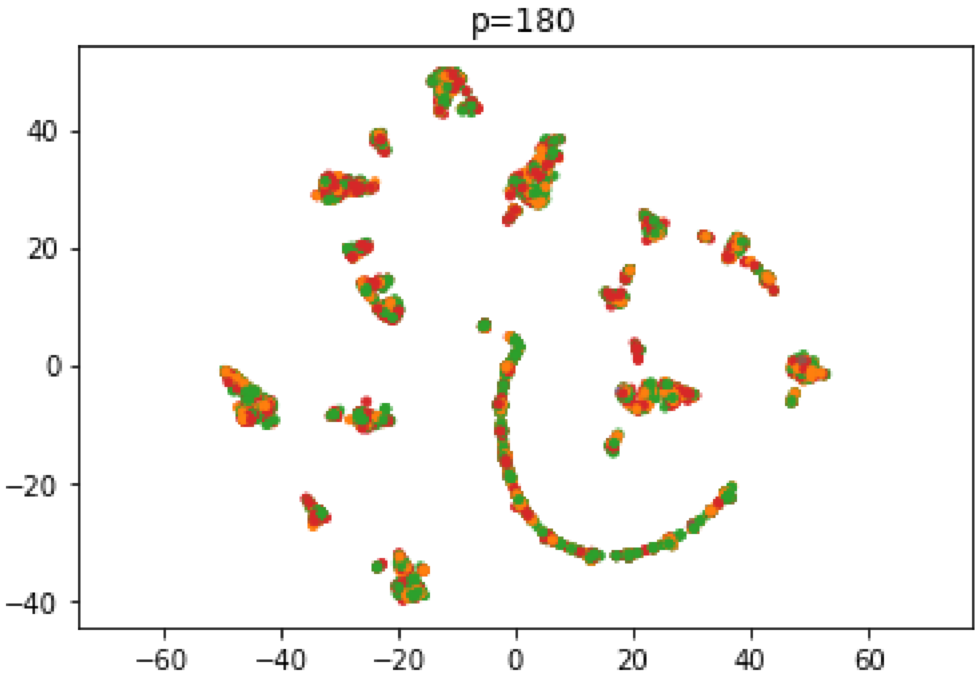

p=130
40
20
0
−20
−40
−60
−75
−50
−25
0
25
50
75

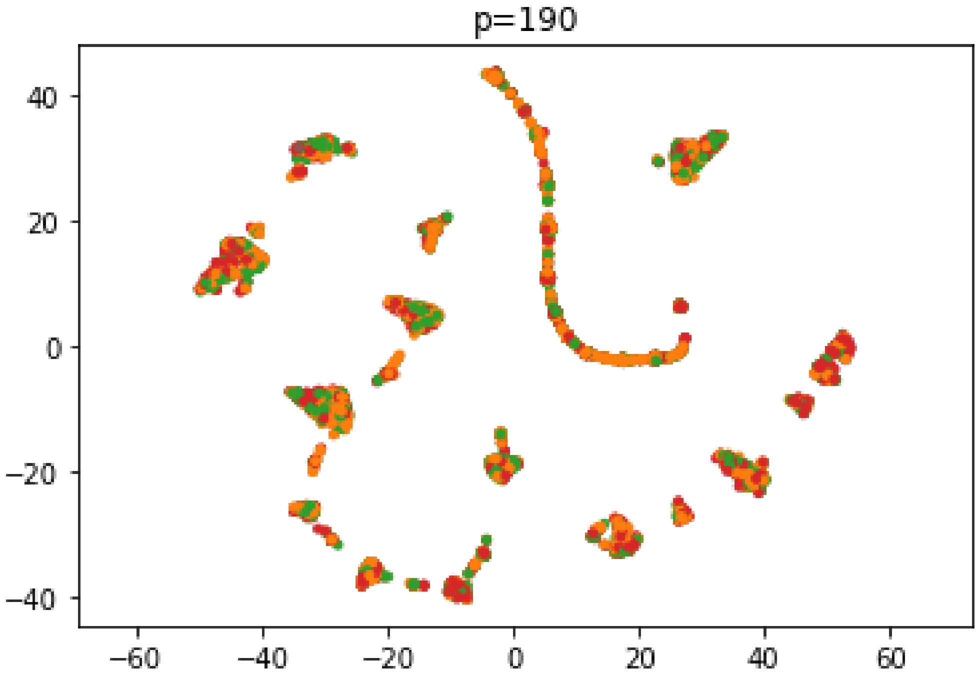

p=140
40
20
0
−20
−40
−60
−80
−60
−40
−20
0
20
40
60
80

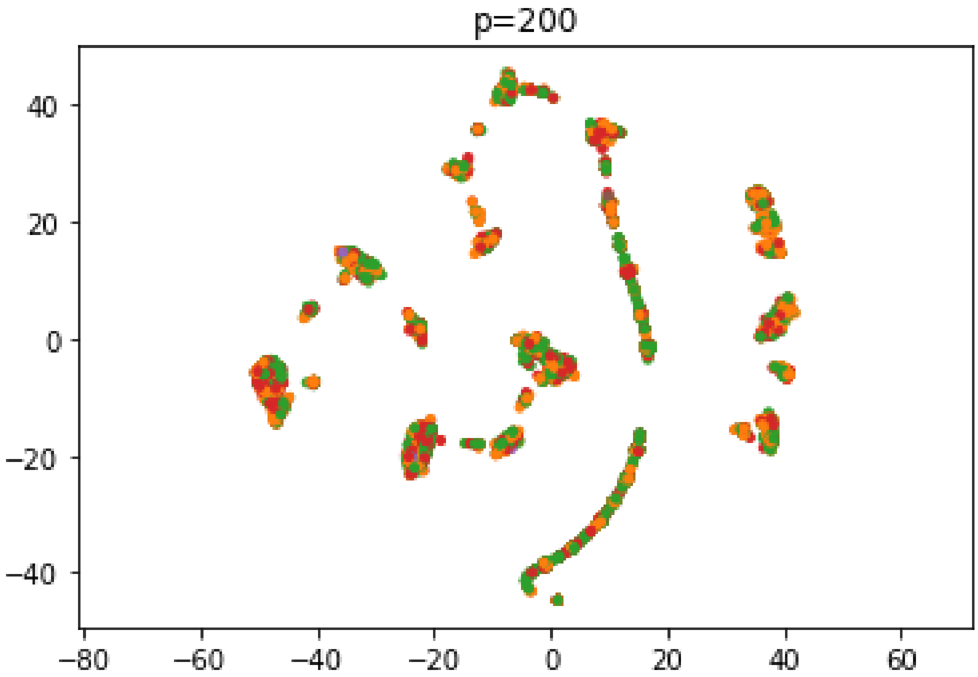

p=150
60
40
20
0
−20
−40
−60
−100
−75
−50
−25
0
25
50
75


p=160
40
20
0
−20
−40
−80
−60
−40
−20
0
20
40
60
80


p=170
40
20
0
−20
−40
−80
−60
−40
−20
0
20
40
60
80


p=180
40
20
0
−20
−40
−60
−40
−20
0
20
40
60

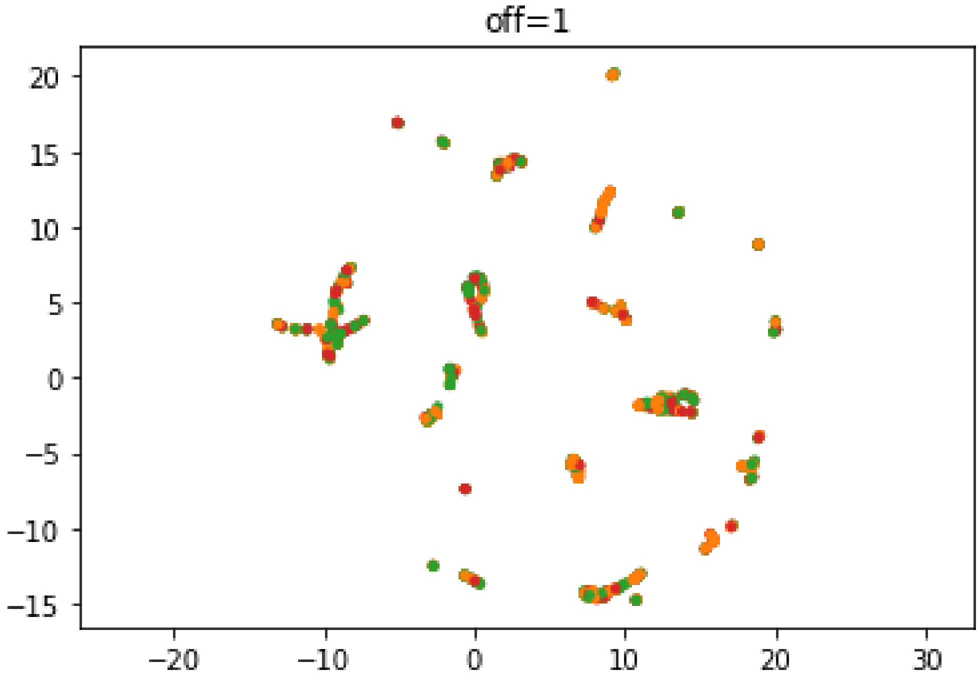

p=190
40
20
0
−20
−40
−60
−40
−20
0
20
40
60



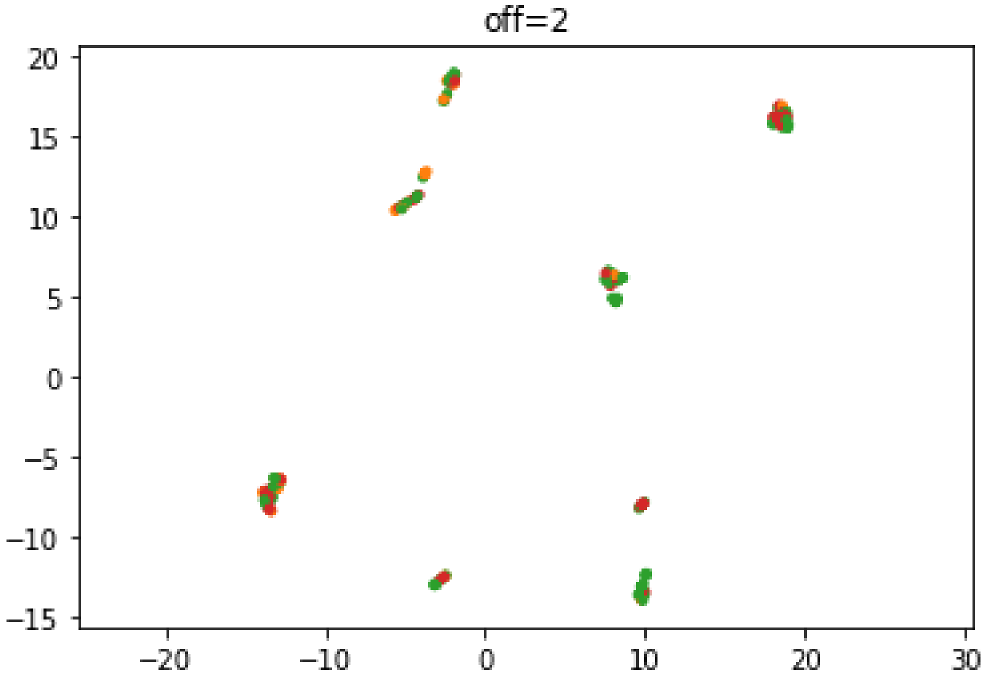
```
In [70]: for off in set(samp.OFFENSE):

             mask = samp_X.OFFENSE == off

             data = samp_X[mask]

         #     trans = TSNE(perplexity=20).fit_transform(data)
             try:
                 trans = reducer.fit_transform(data.values)

                 plt.scatter(trans[:, 0], trans[:, 1], c=[sns.color_palette()[x] for x in samp_
                 plt.gca().set_aspect('equal', 'datalim')
                 plt.title(f'off={off}')
                 plt.show()
             except:

                 continue
```

```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 5 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Us
```

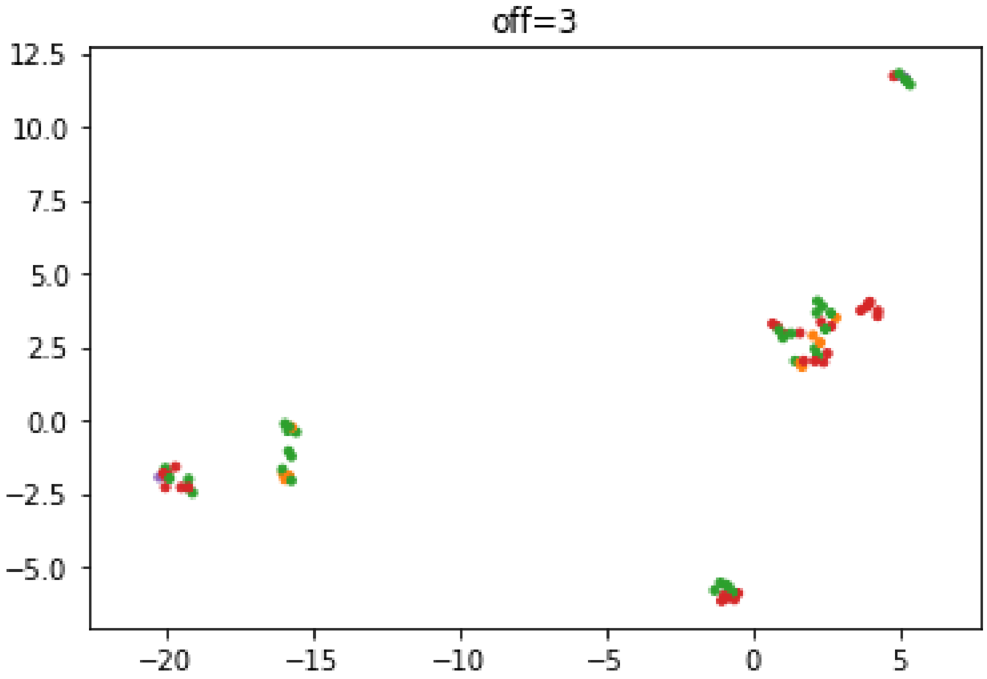
```
Graph is not fully connected, spectral embedding may not work as expected.
```



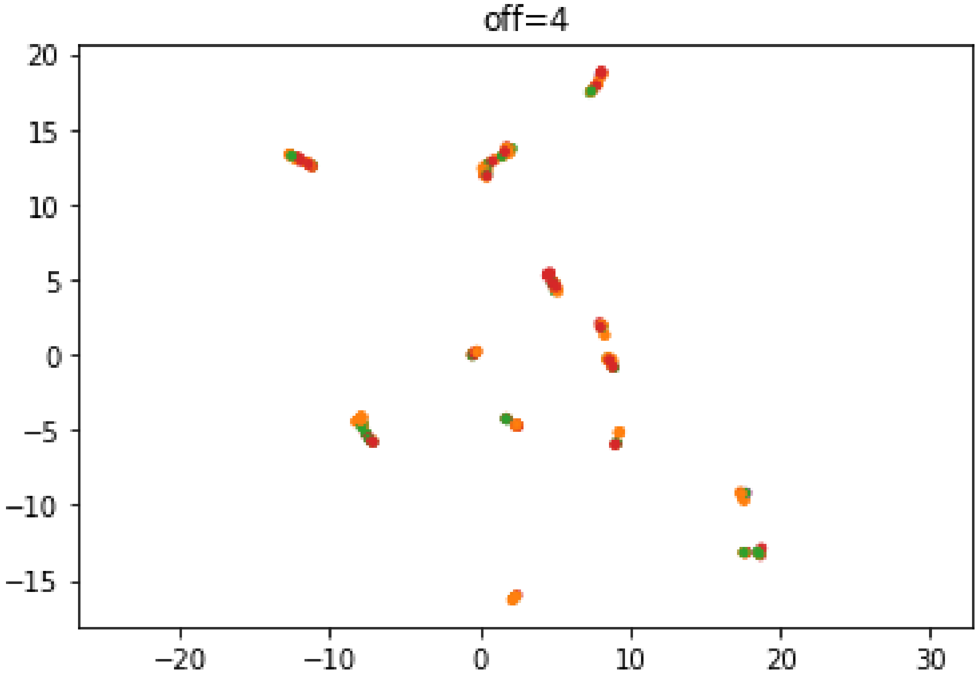
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 4 separate connected components using meta-embedding (experimental)
```

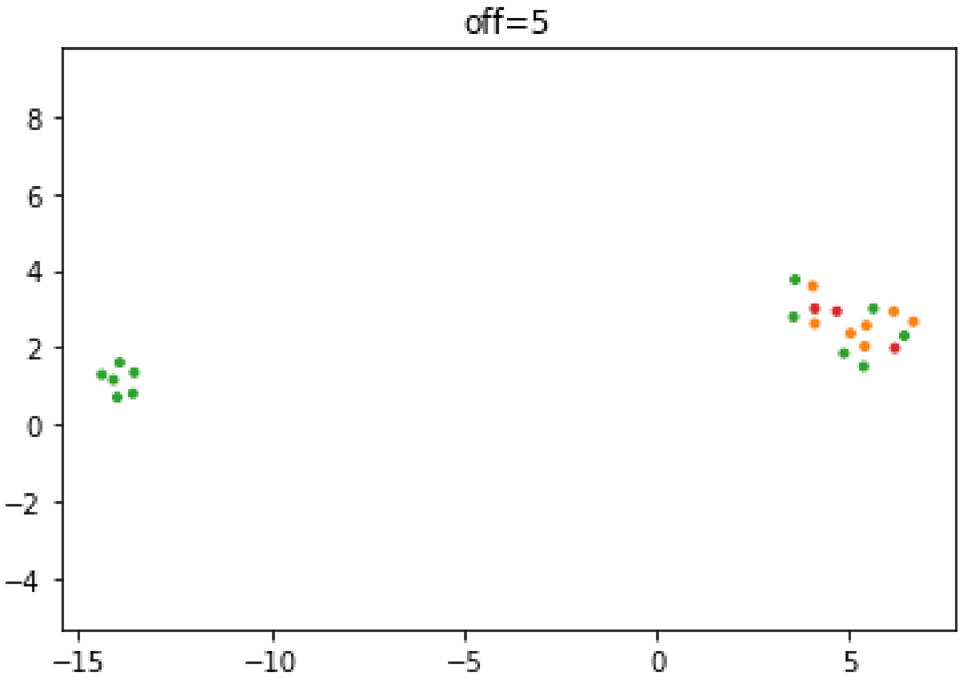


```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 2 separate connected components using meta-embedding (experimental)
```

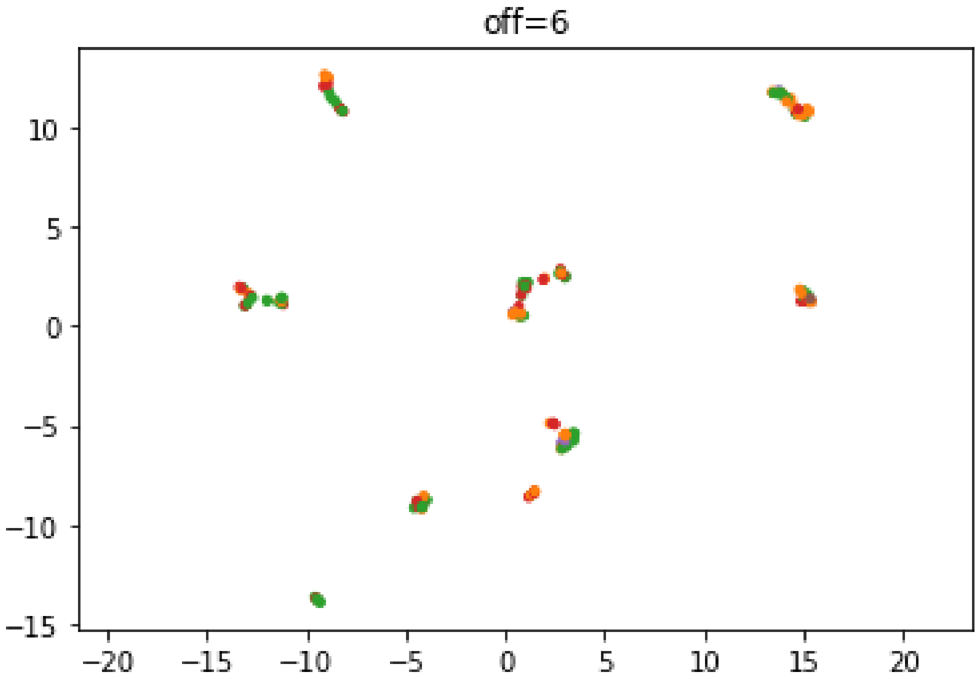


```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 4 separate connected components using meta-embedding (experimental)
```

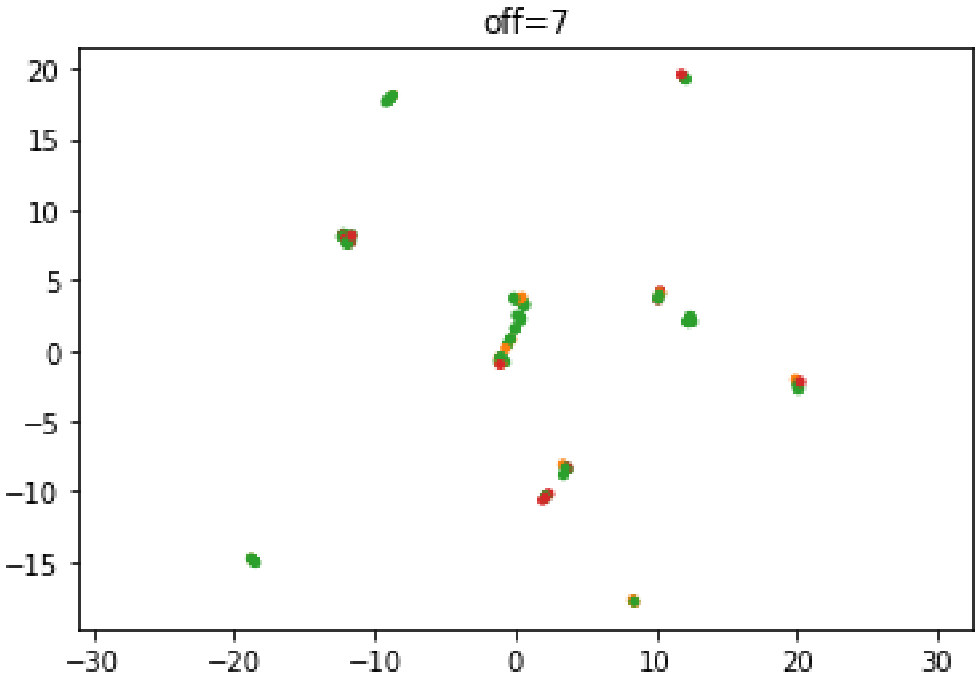




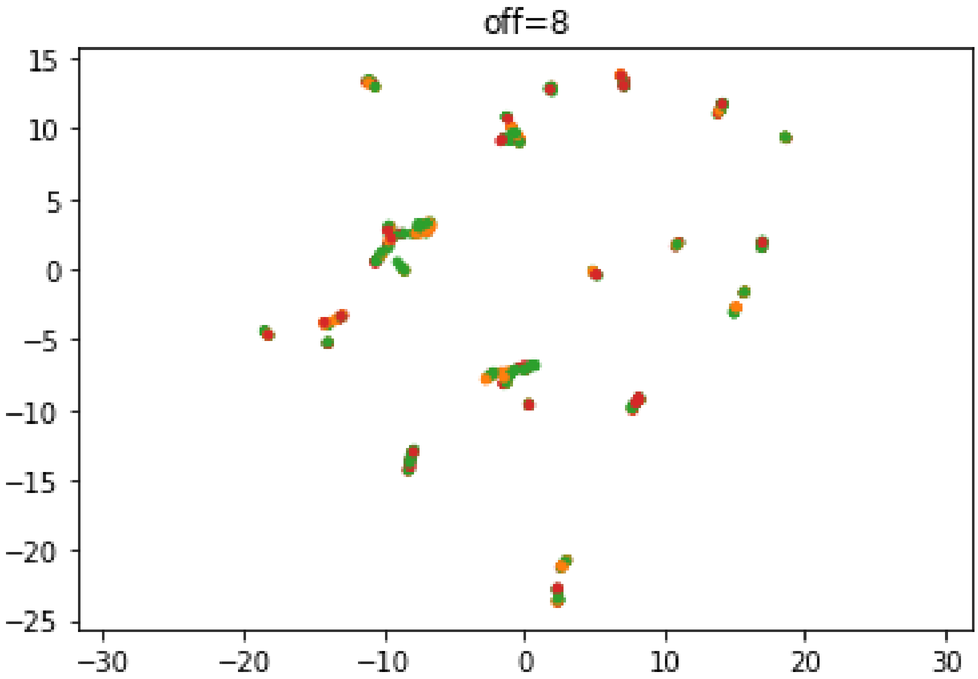
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 6 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```



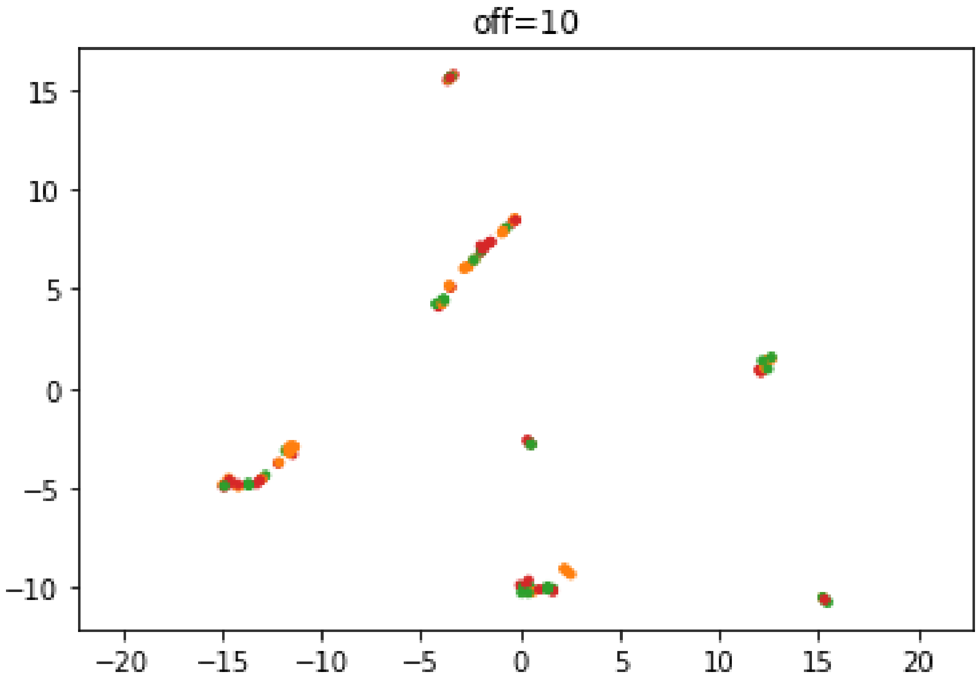
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 8 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Us

Graph is not fully connected, spectral embedding may not work as expected.
```



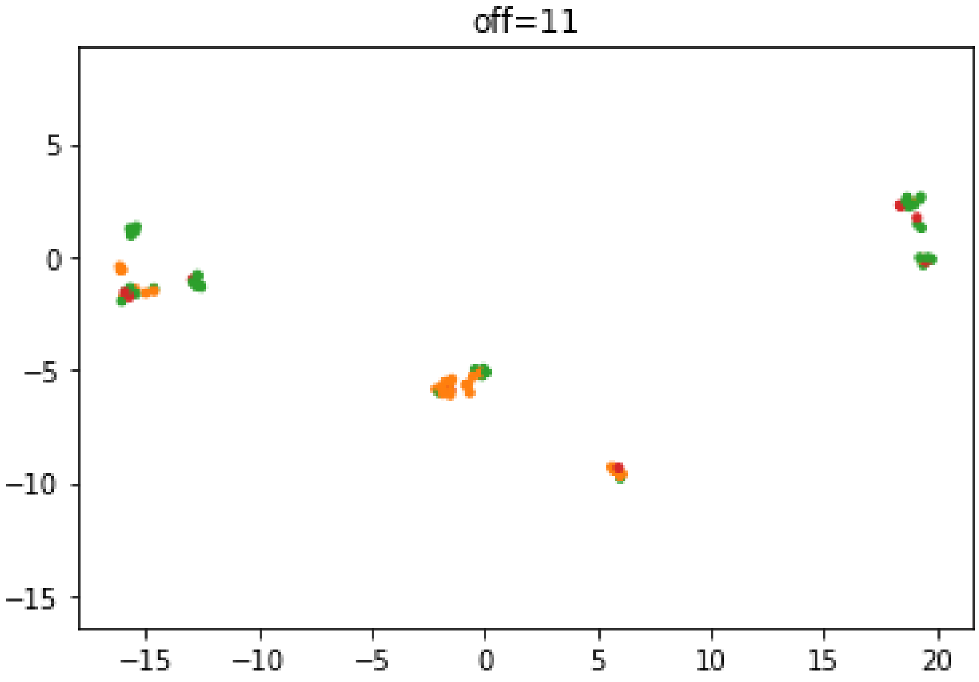
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 9 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Us

Graph is not fully connected, spectral embedding may not work as expected.
```



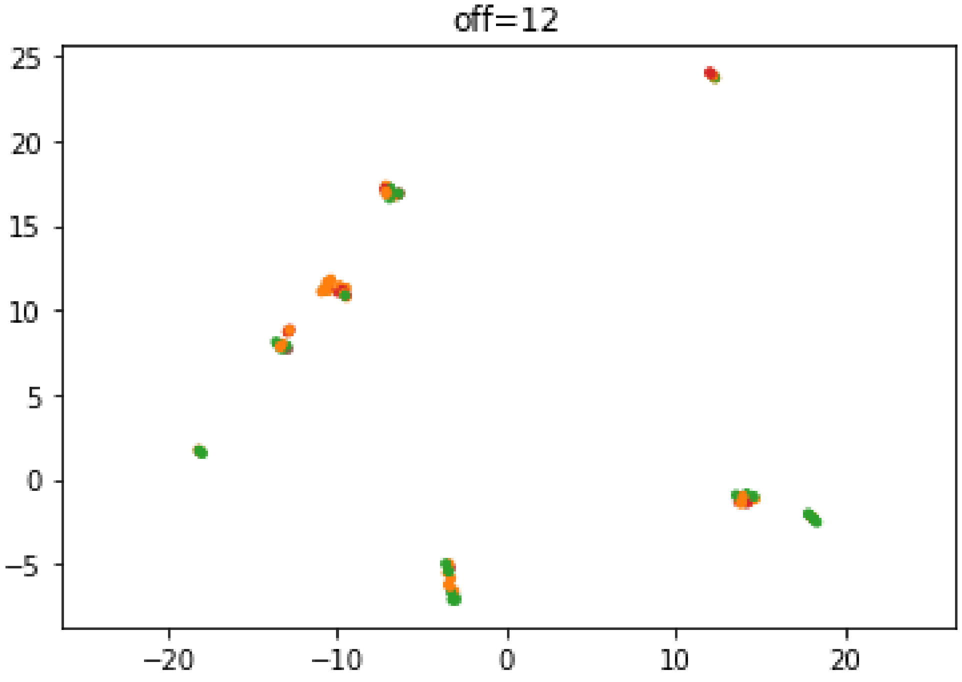
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592

k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.

/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 3 separate connected components using meta-embedding (experimental)
```



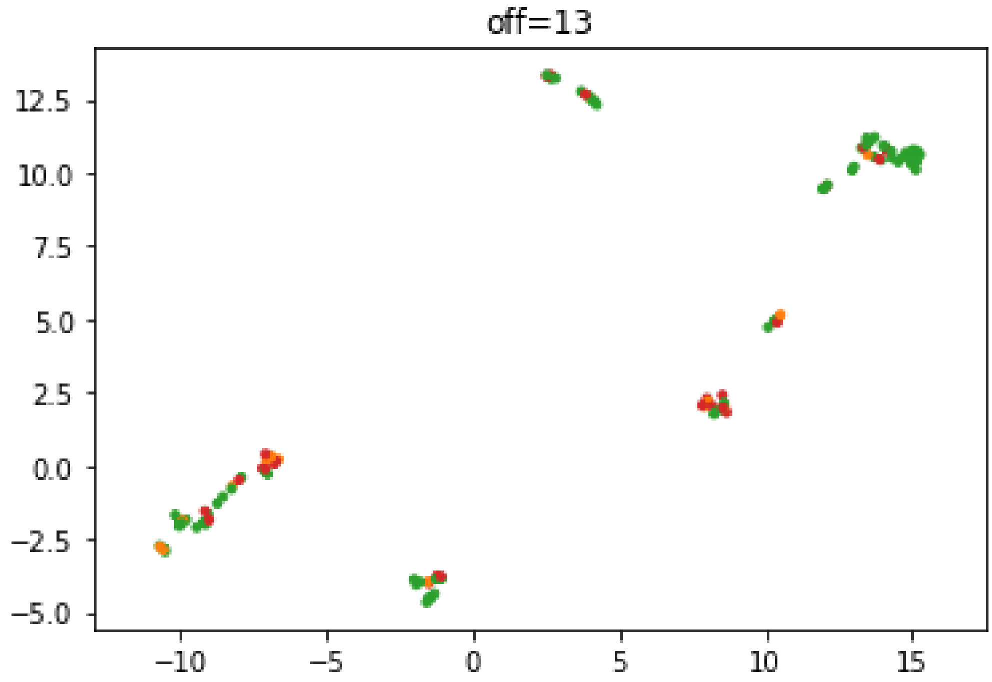
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 2 separate connected components using meta-embedding (experimental)
```



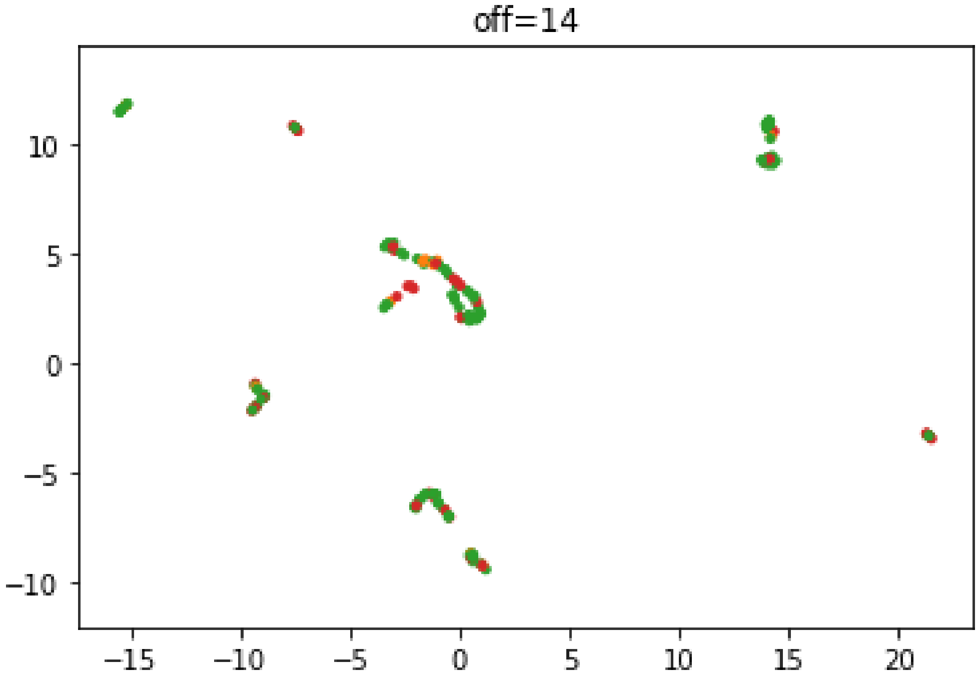
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 5 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Us

Graph is not fully connected, spectral embedding may not work as expected.
```



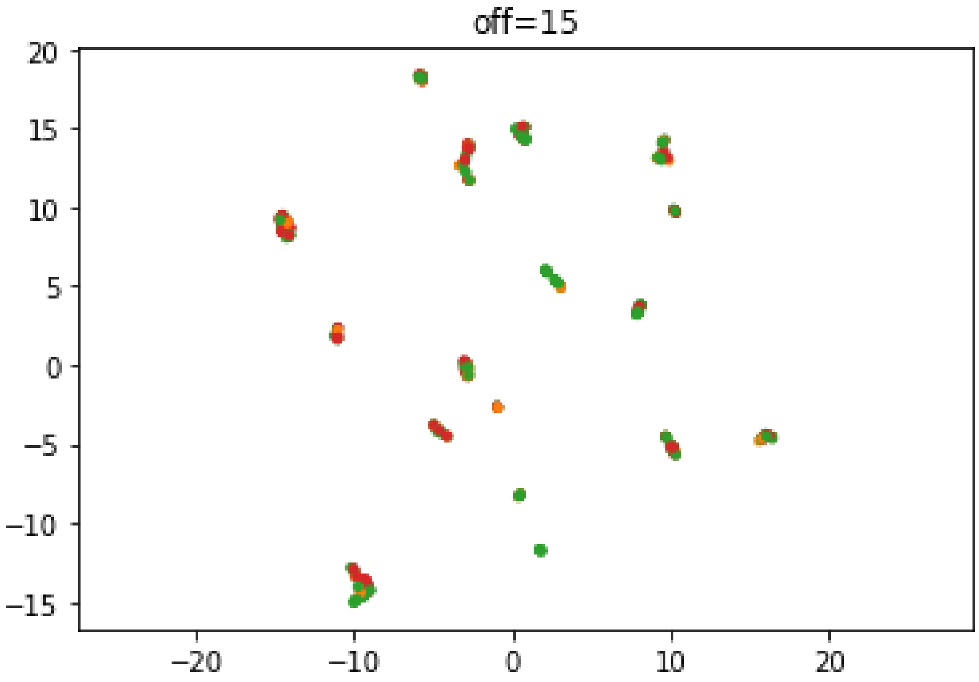
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 2 separate connected components using meta-embedding (experimental)
```



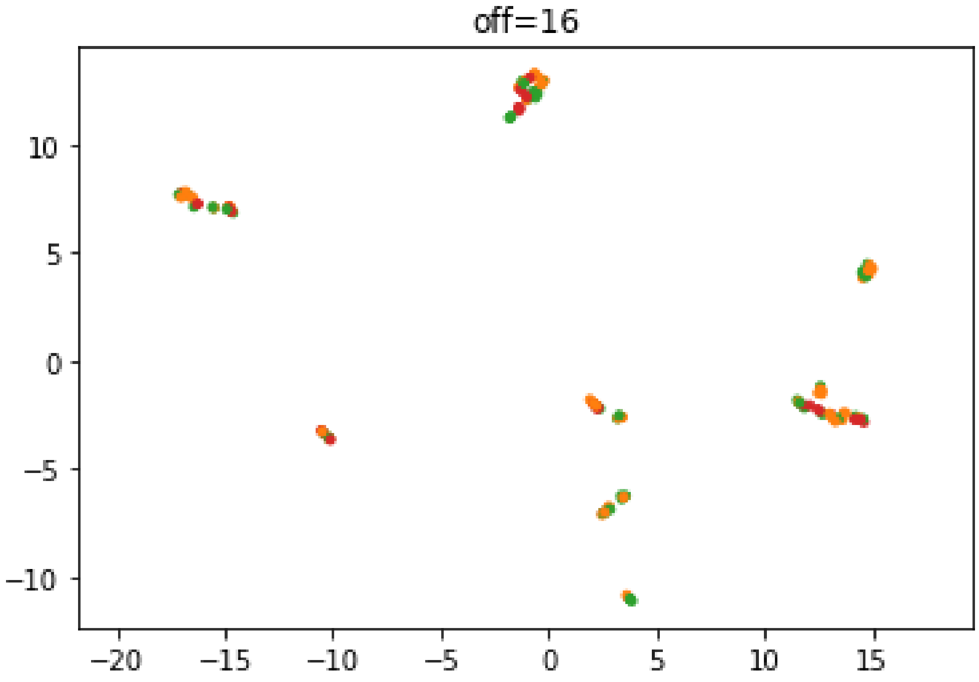
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 5 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```

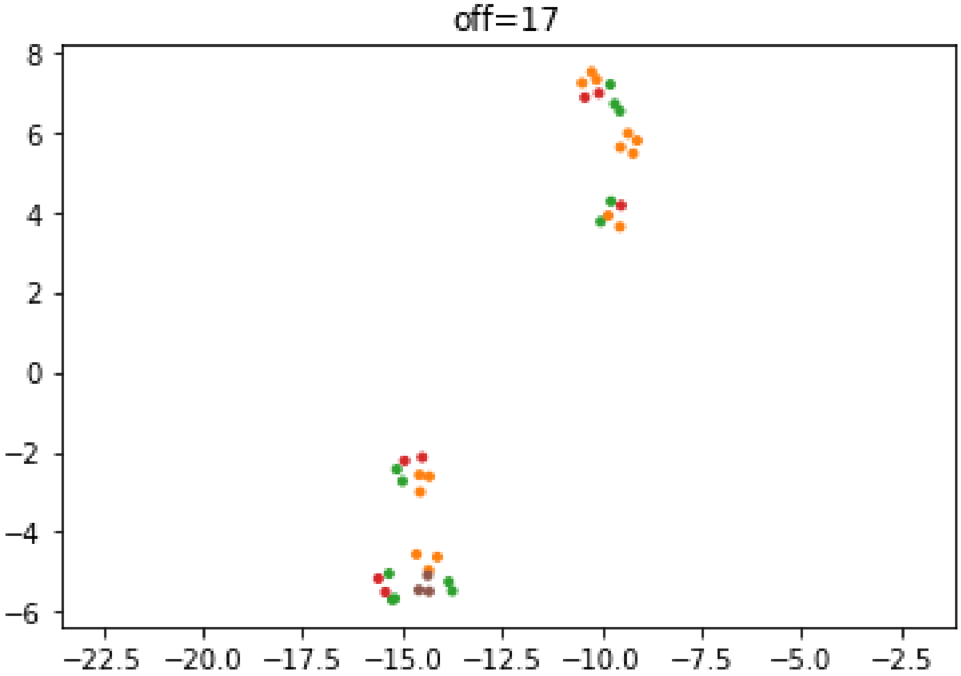


```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 7 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```

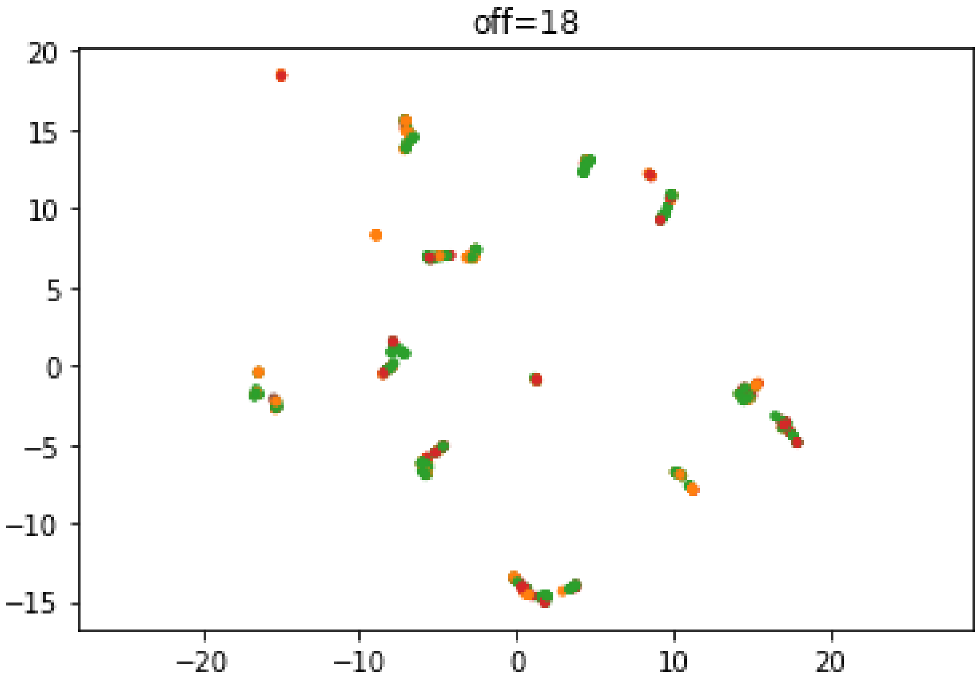


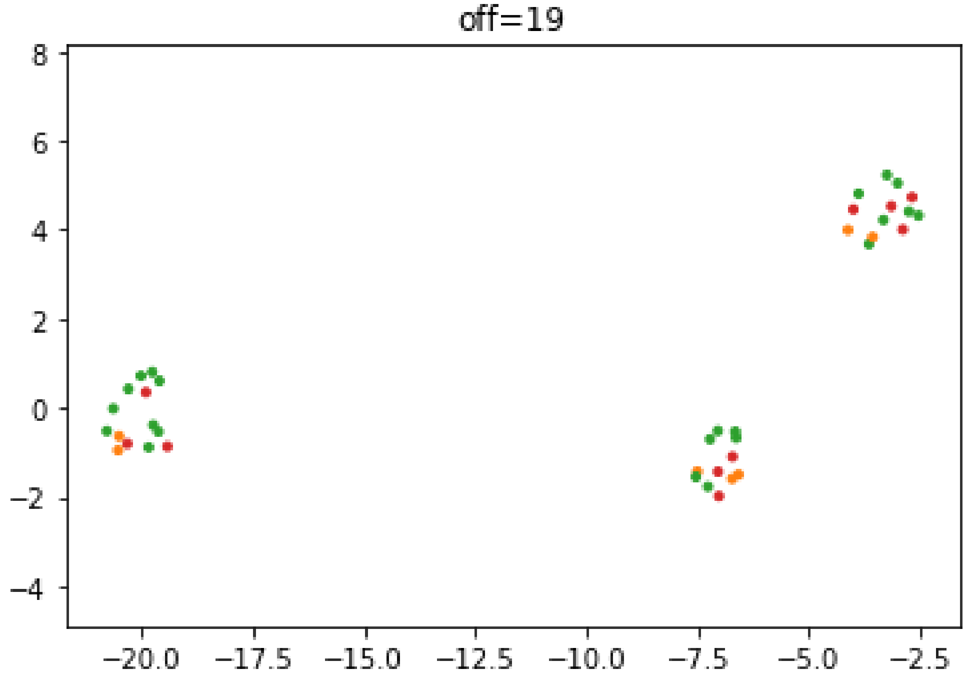
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 3 separate connected components using meta-embedding (experimental)
```





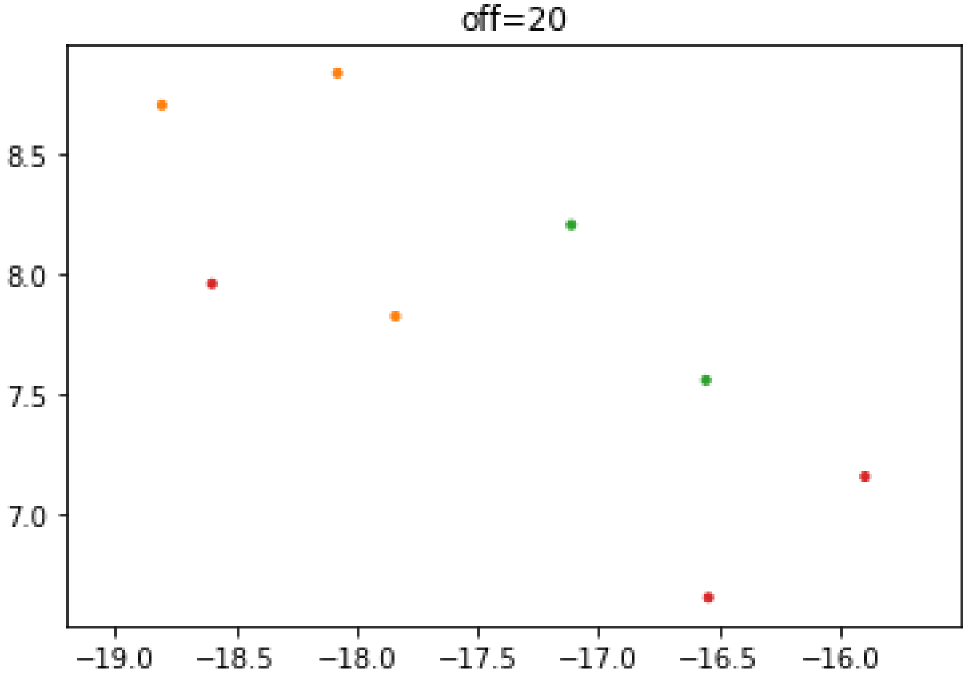
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 8 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```

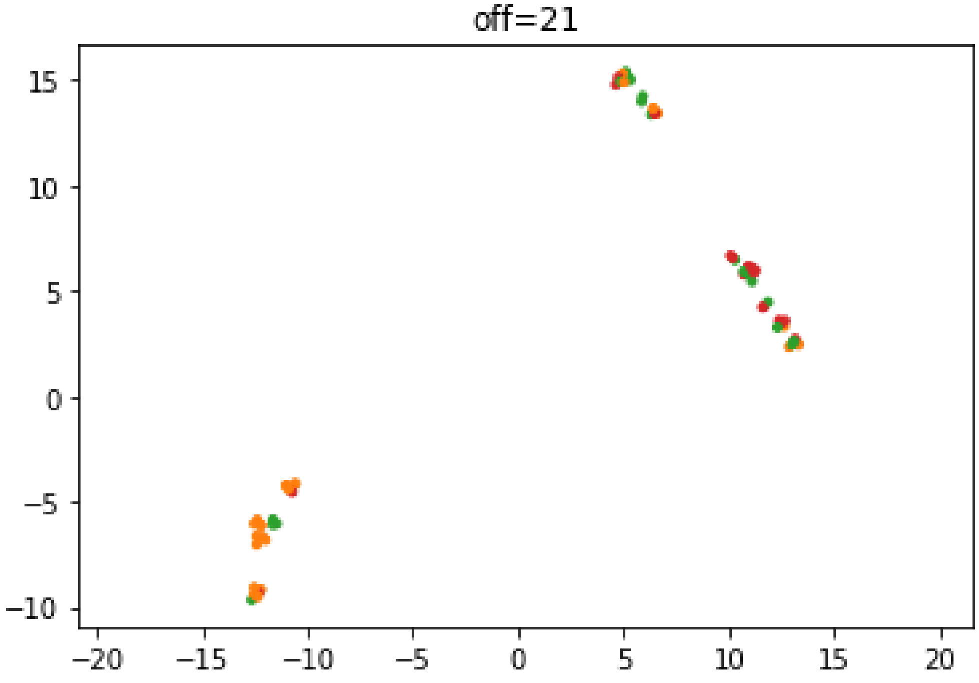


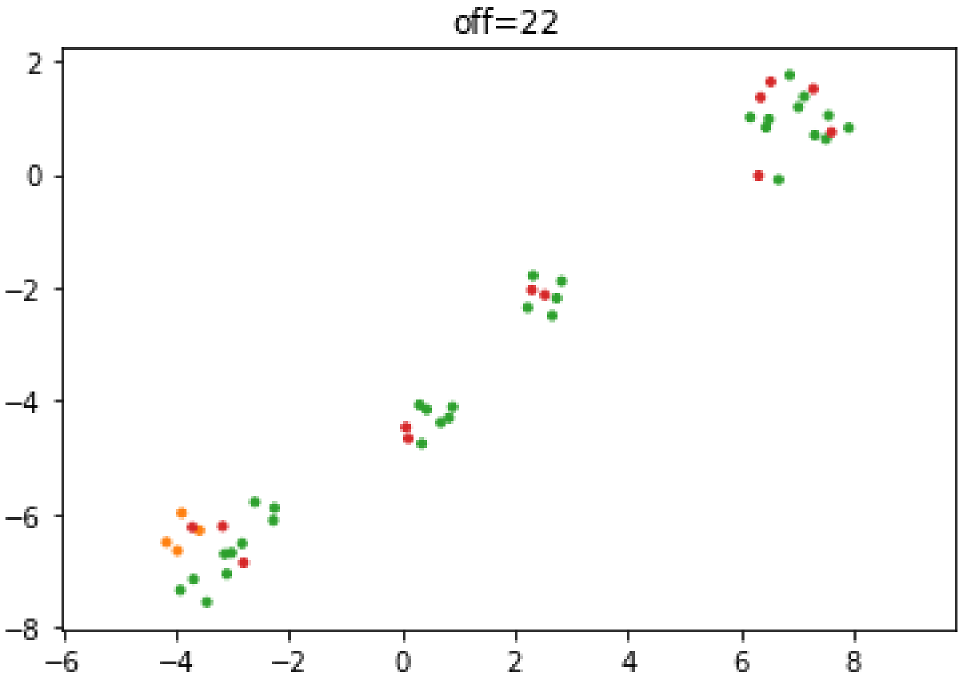


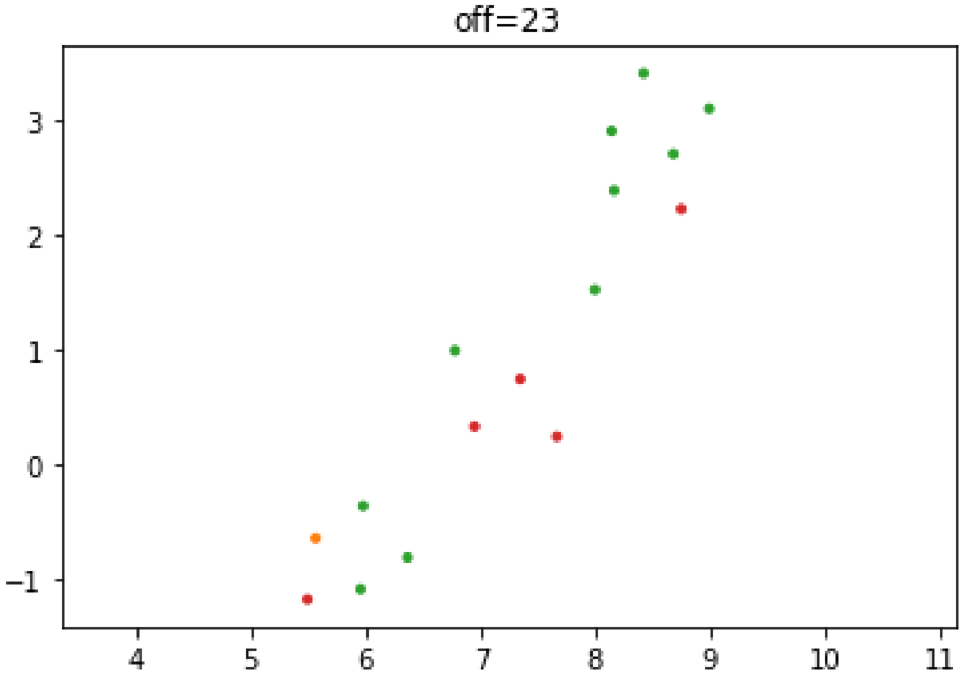
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



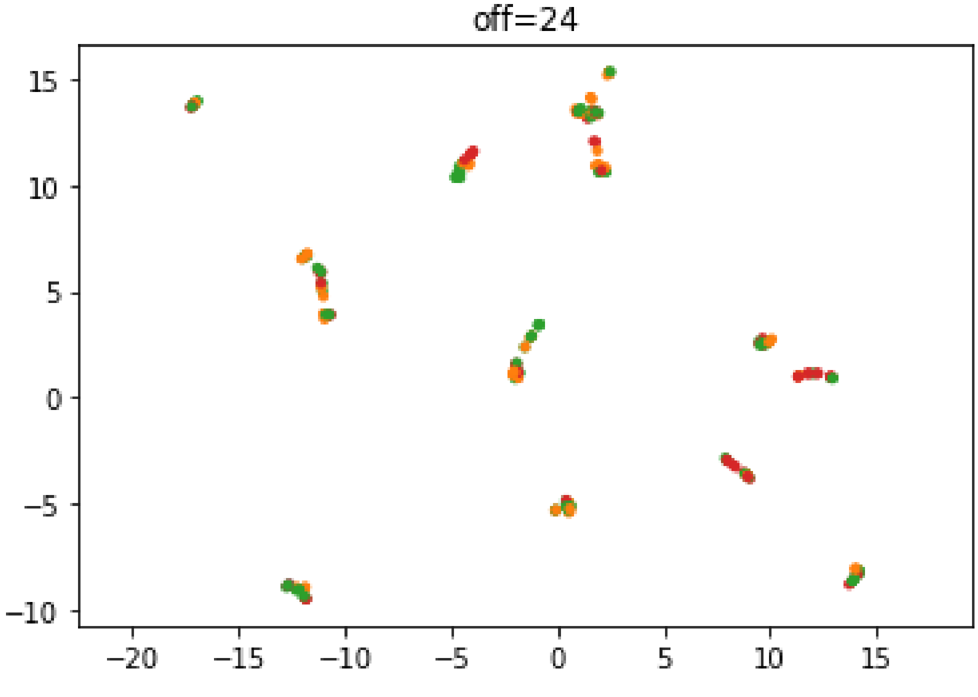




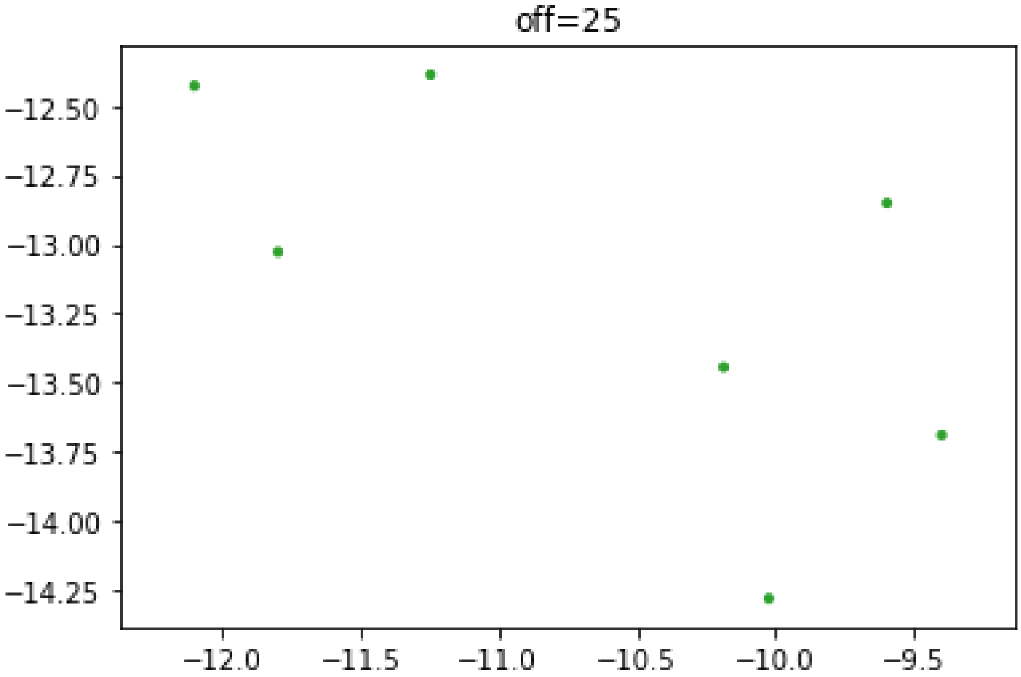


```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:
```

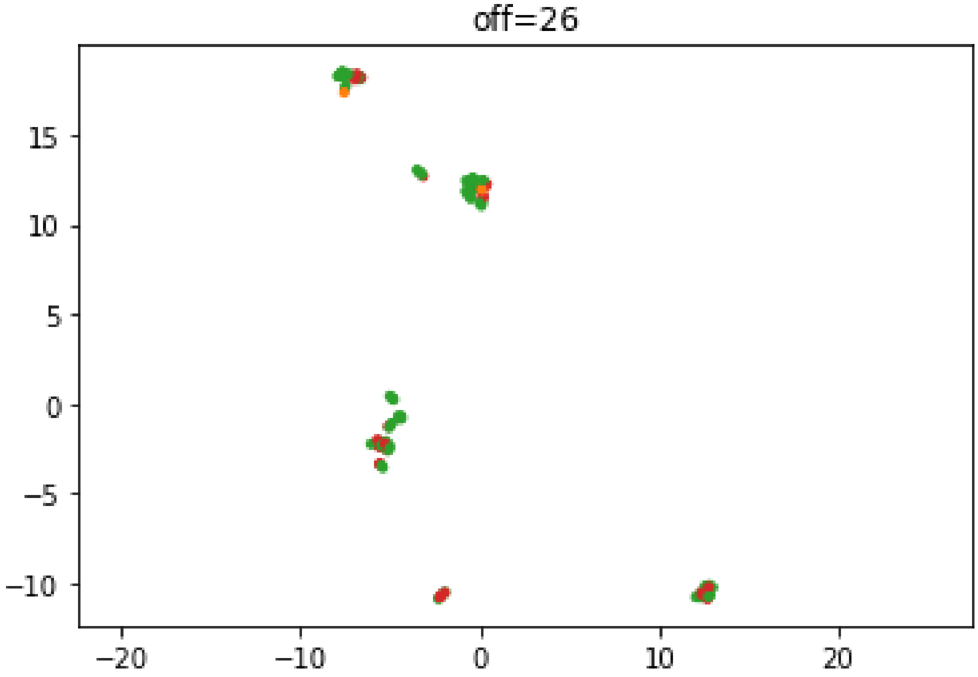
```
Embedding a total of 4 separate connected components using meta-embedding (experimental)
```



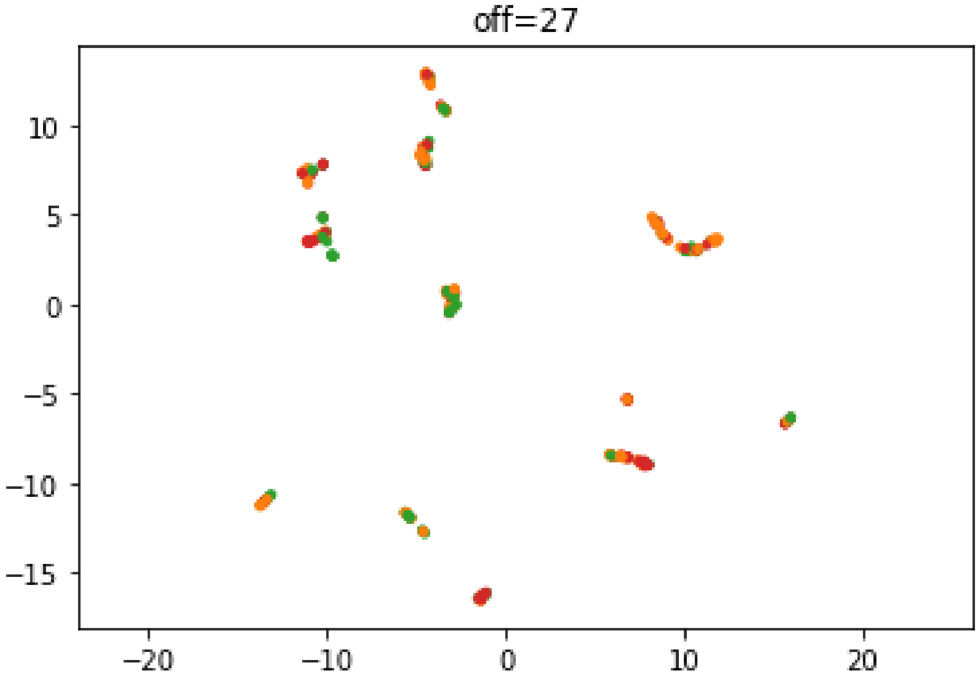
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

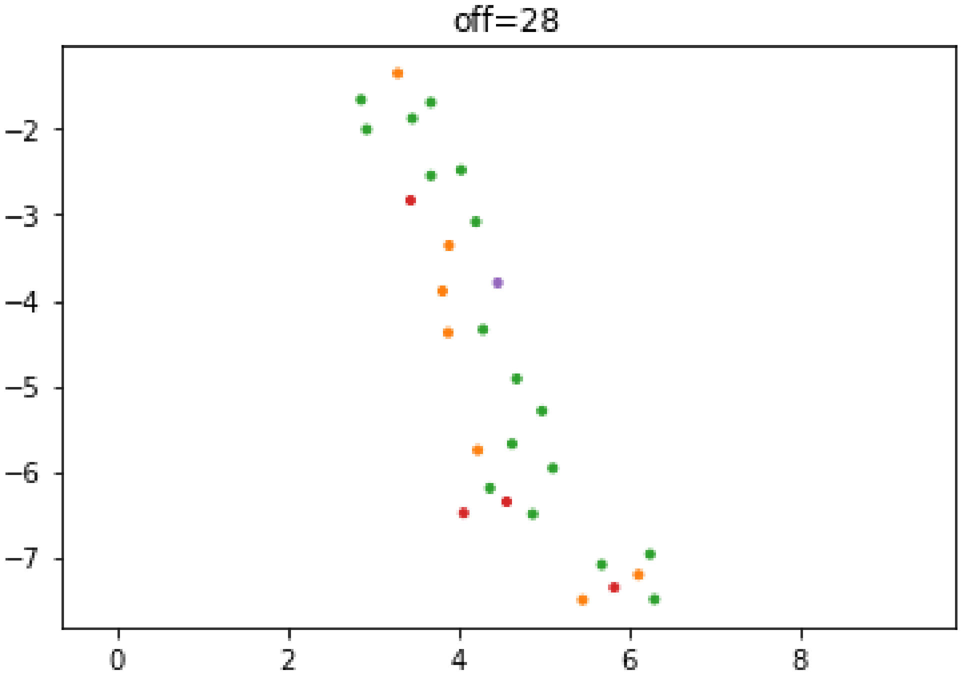


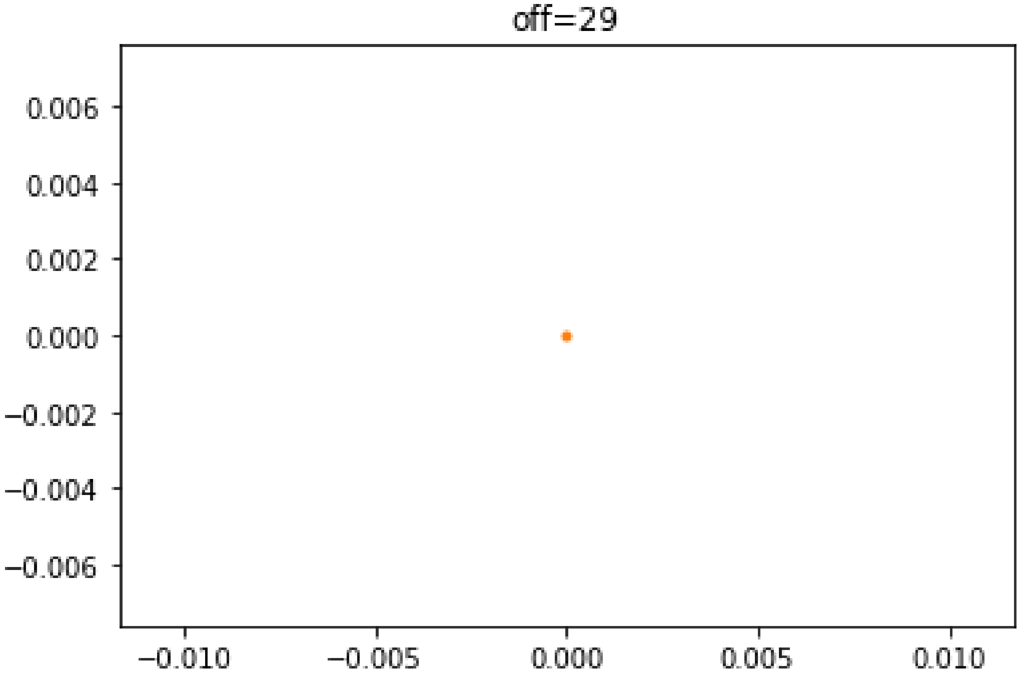
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 4 separate connected components using meta-embedding (experimental)
```

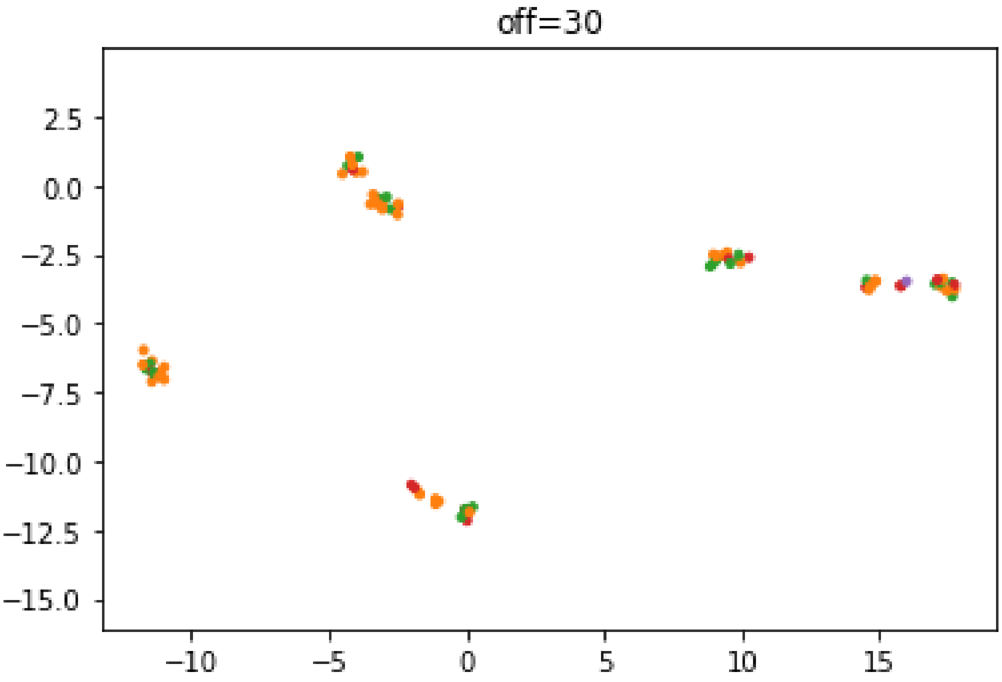


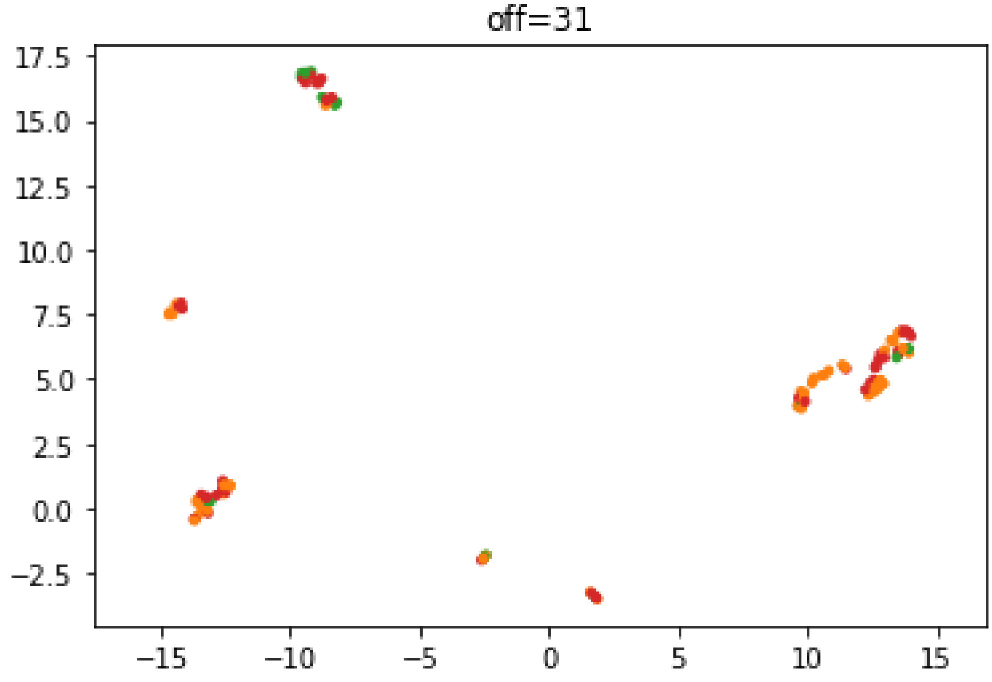
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 4 separate connected components using meta-embedding (experimental)
```

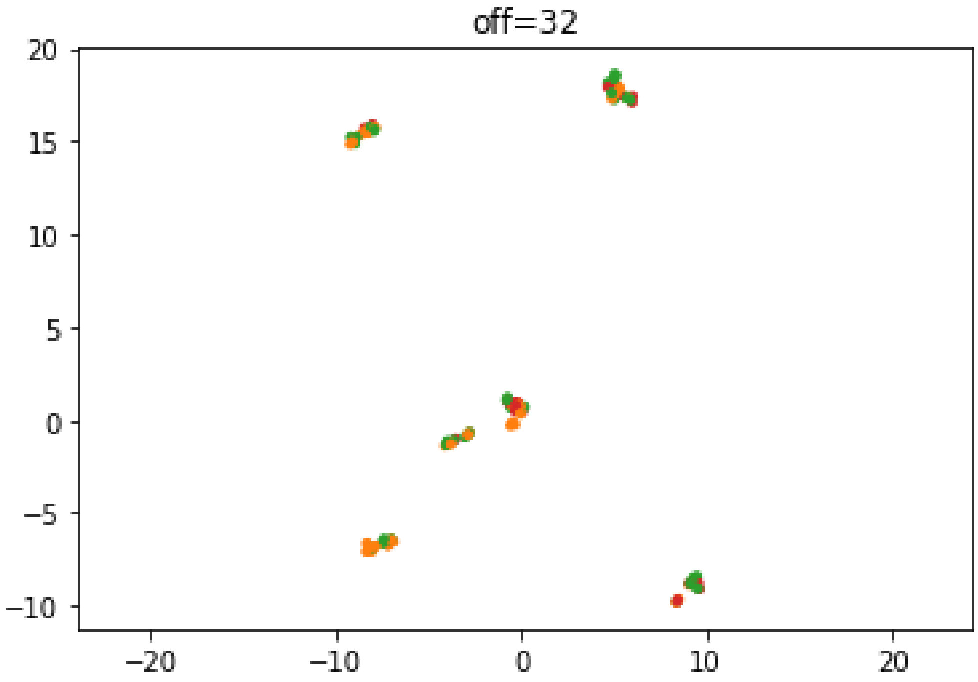


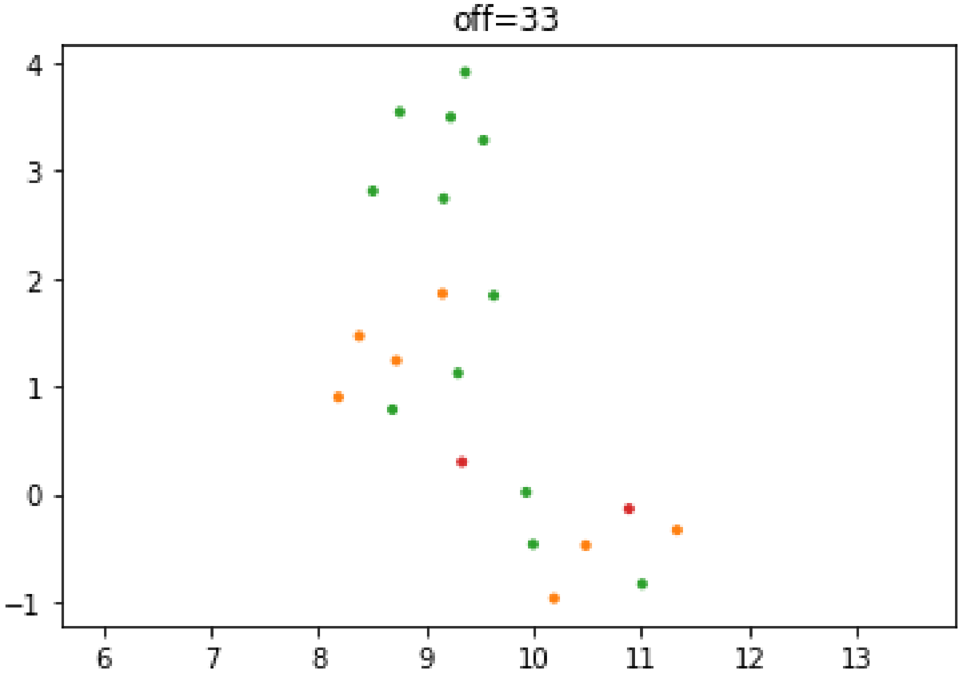

off=28
−2
−3
−4
−5
−6
−7
0
2
4
6
8

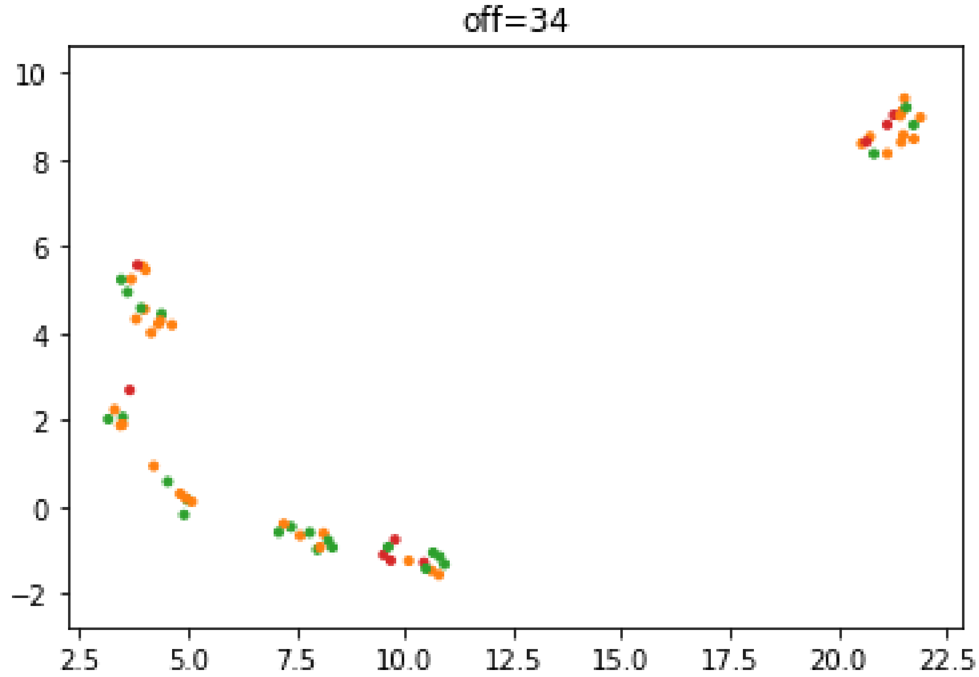

off=29
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010


off=30
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−12.5
−15.0
−10
−5
0
5
10
15

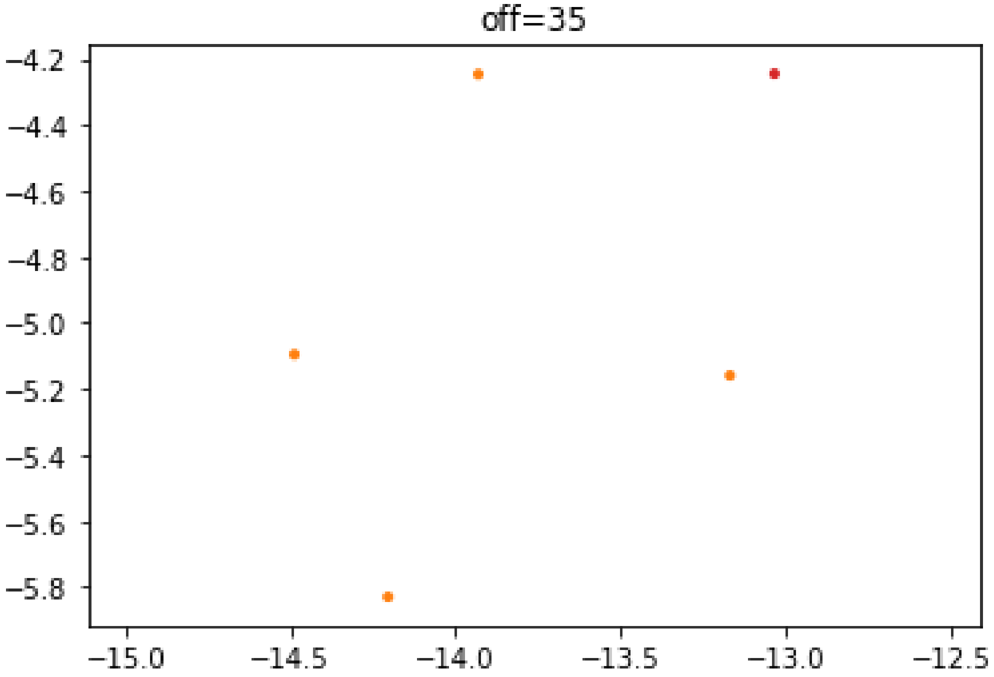

off=31
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−15
−10
−5
0
5
10
15

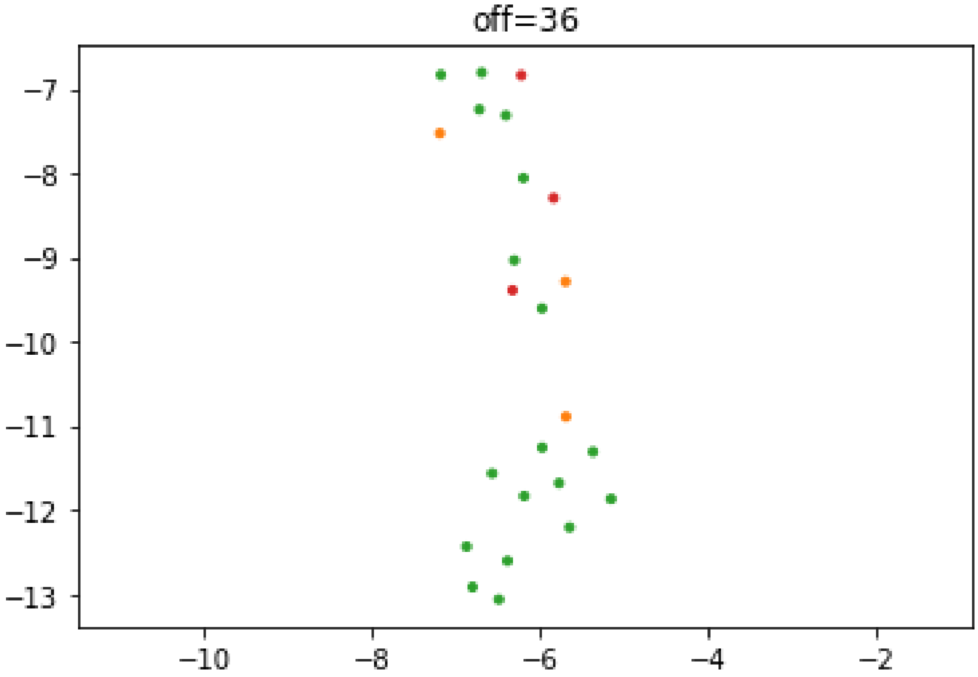

off=32
20
15
10
5
0
−5
−10
−20
−10
0
10
20

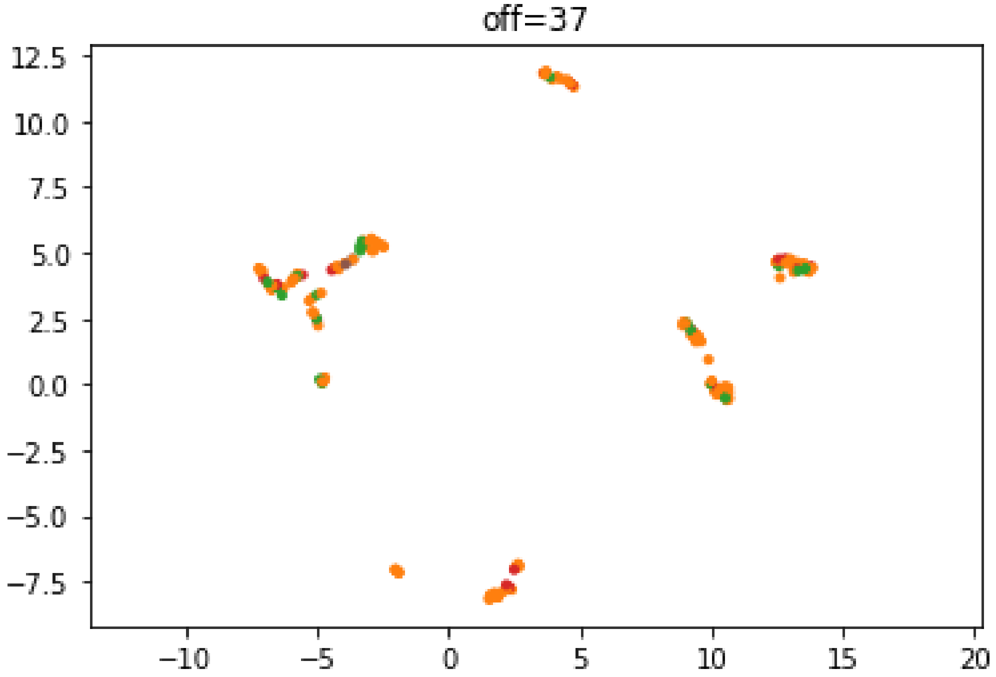

off=33
4
3
2
1
0
−1
6
7
8
9
10
11
12
13



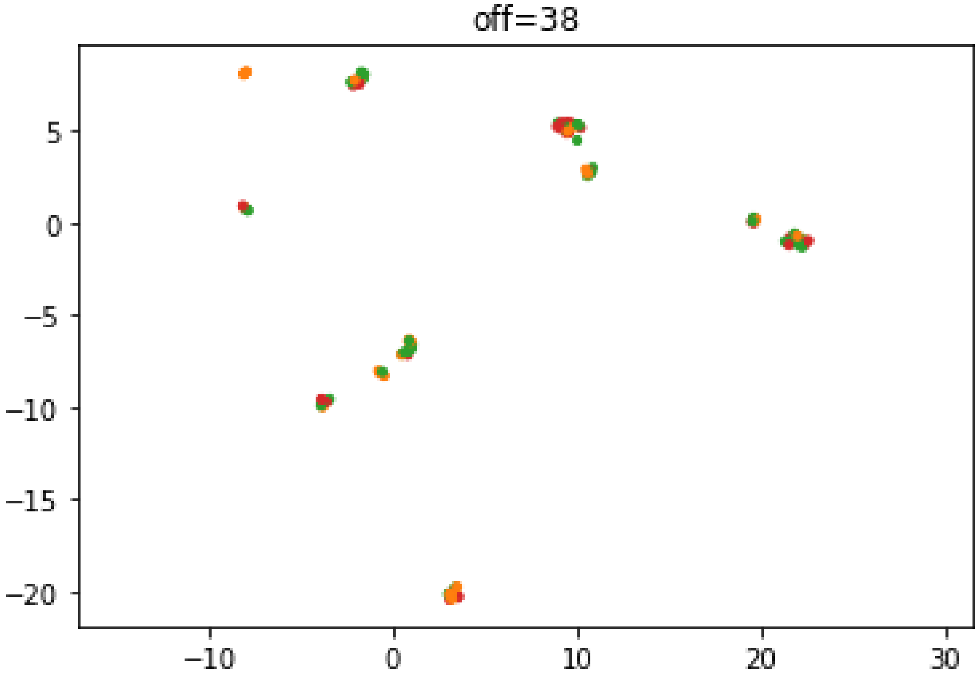
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

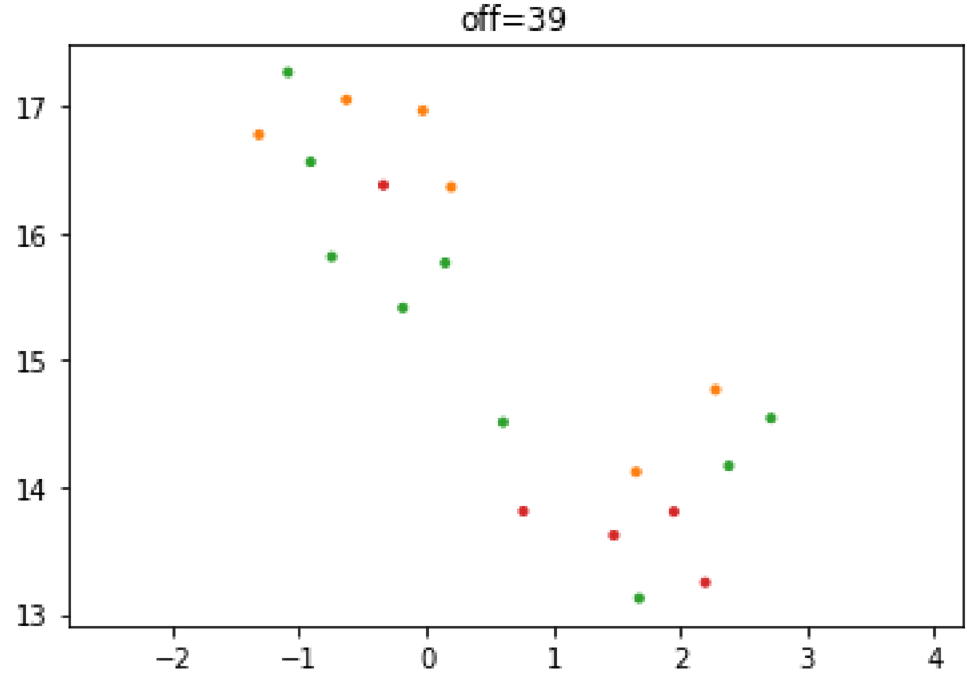


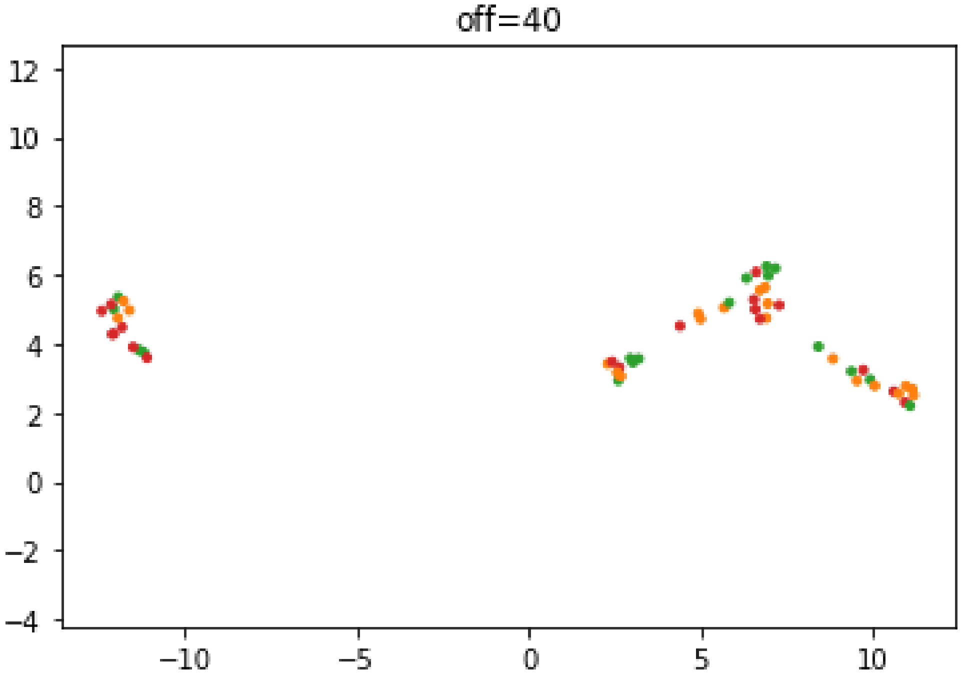

off=36
−7
−8
−9
−10
−11
−12
−13
−10
−8
−6
−4
−2

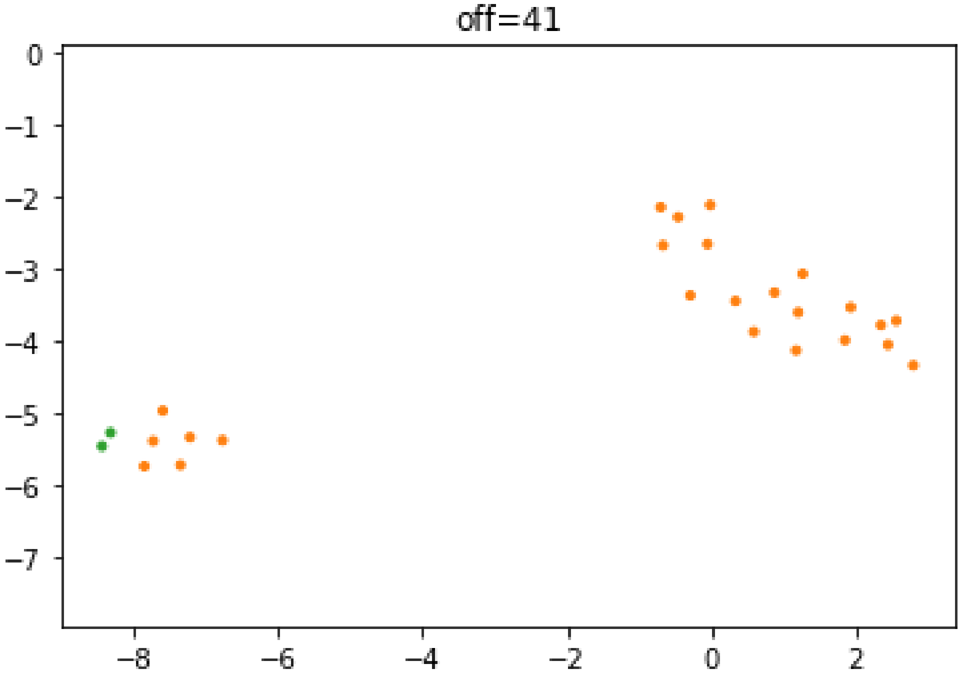

off=37
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10
−5
0
5
10
15
20

```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 2 separate connected components using meta-embedding (experimental)
```

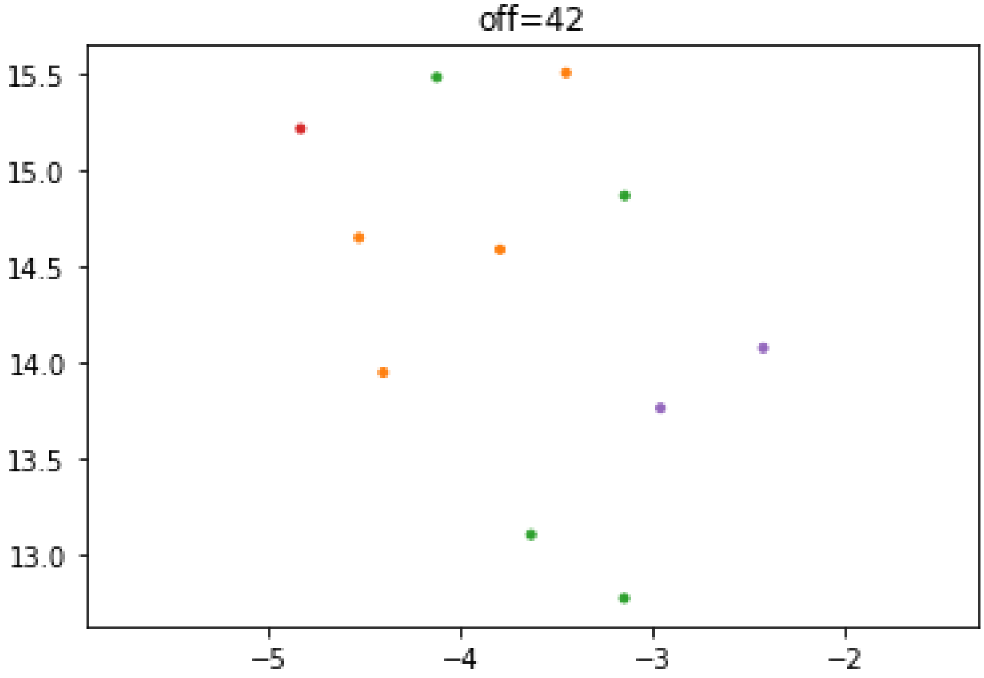


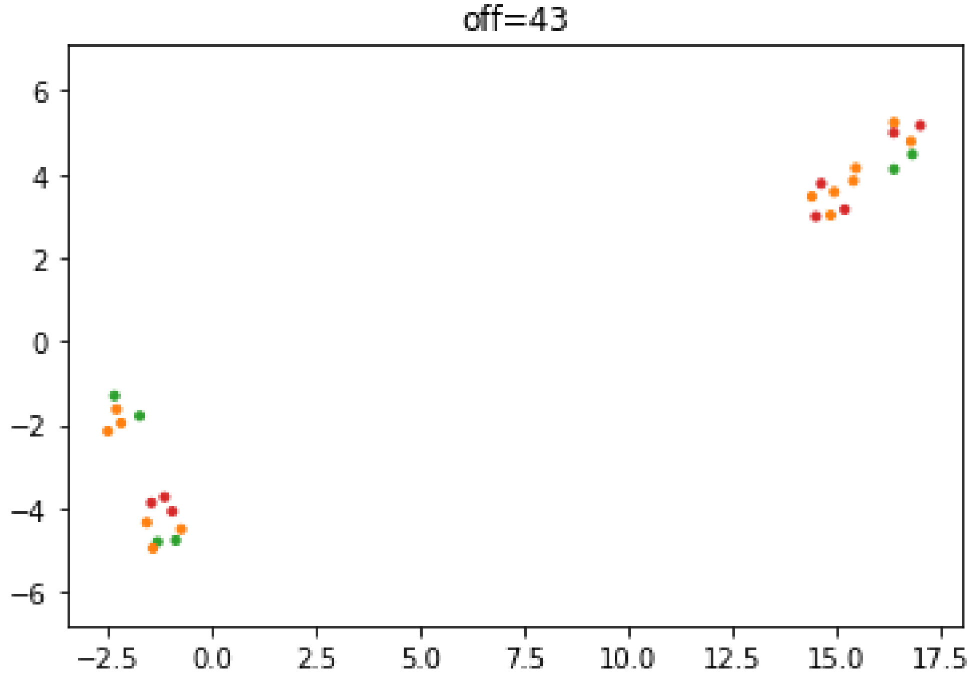



off=40
12
10
8
6
4
2
0
−2
−4
−10
−5
0
5
10

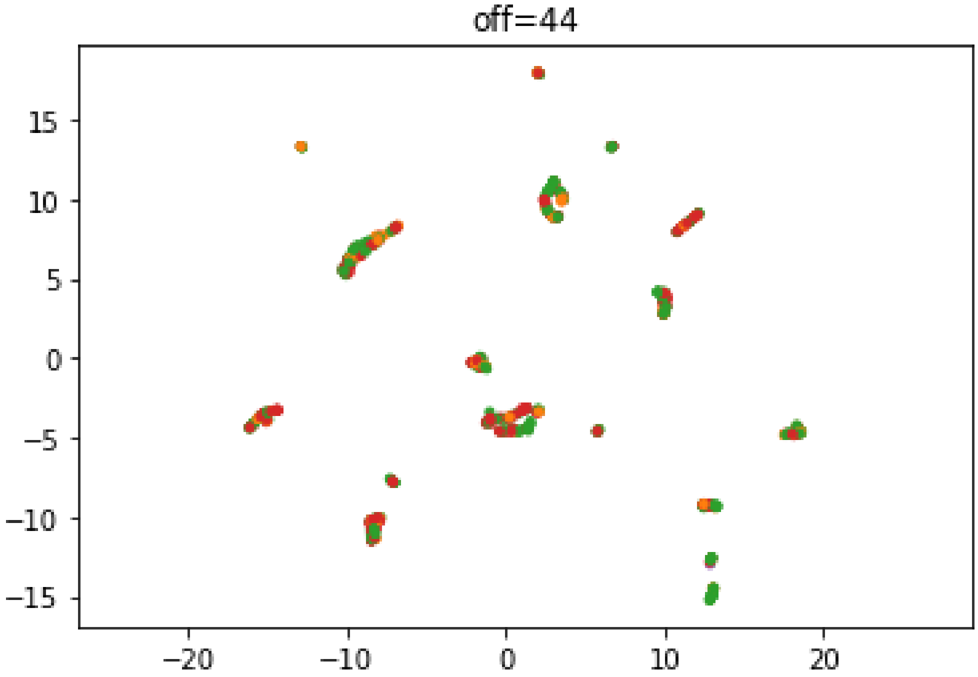

off=41
0
−1
−2
−3
−4
−5
−6
−7
−8
−6
−4
−2
0
2

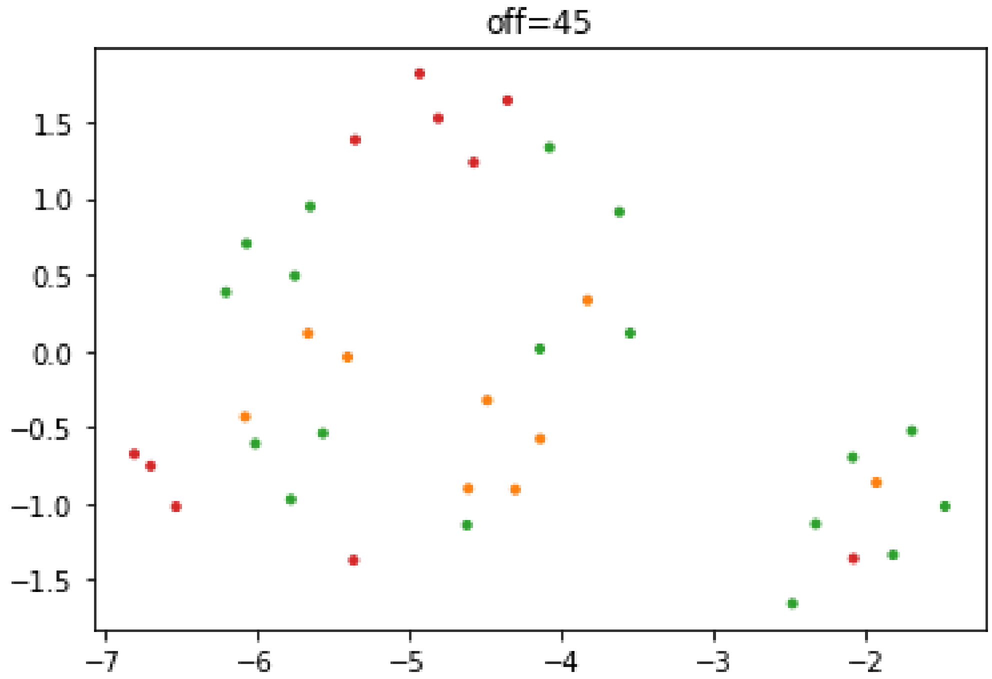
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



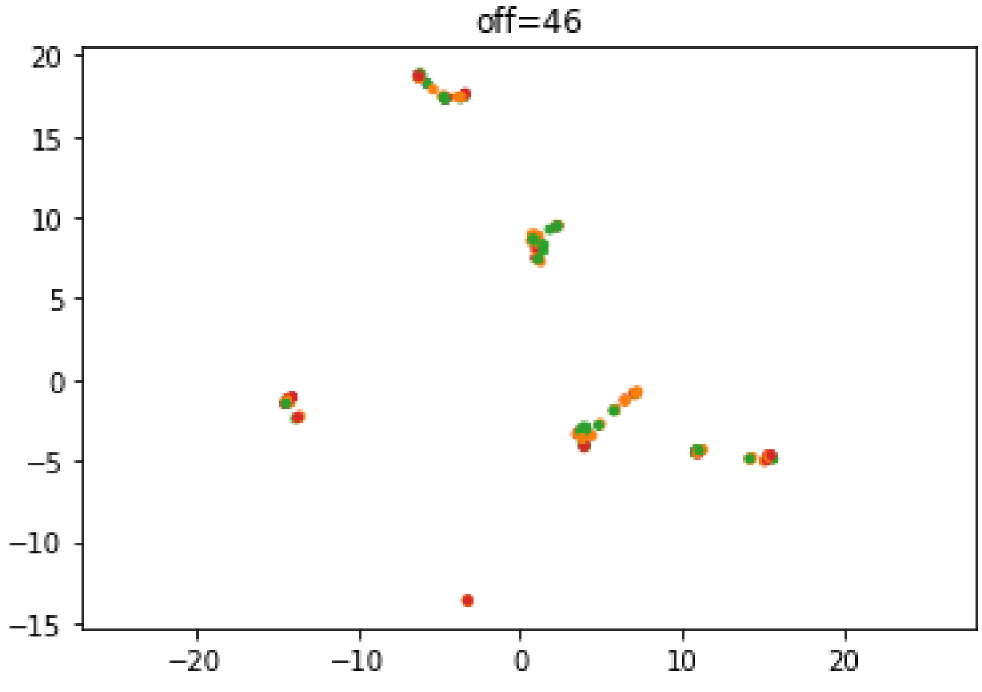


```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 8 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```

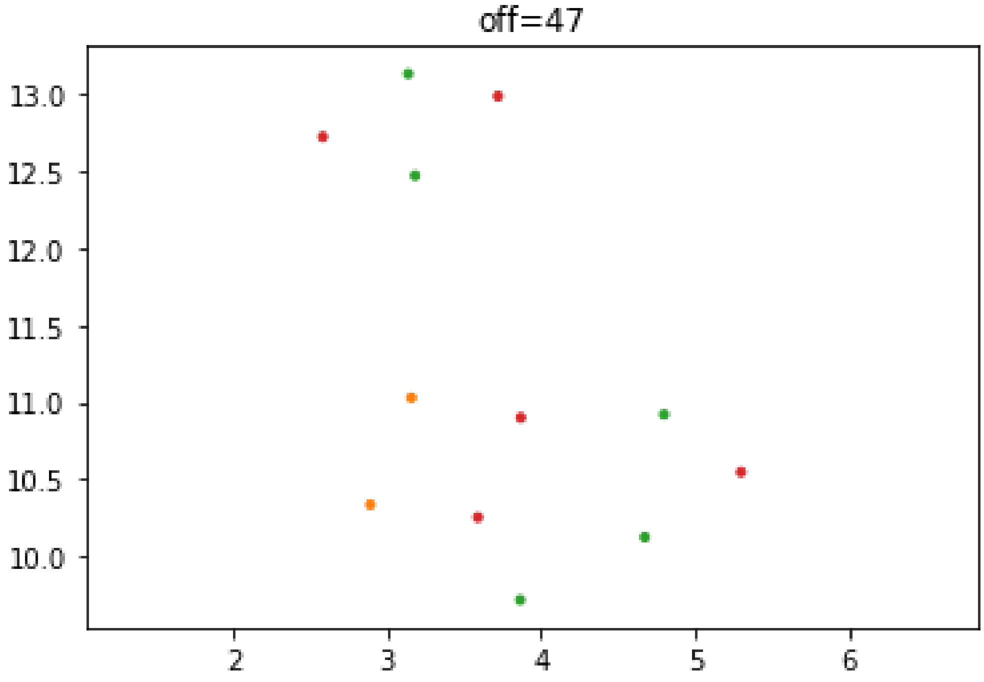




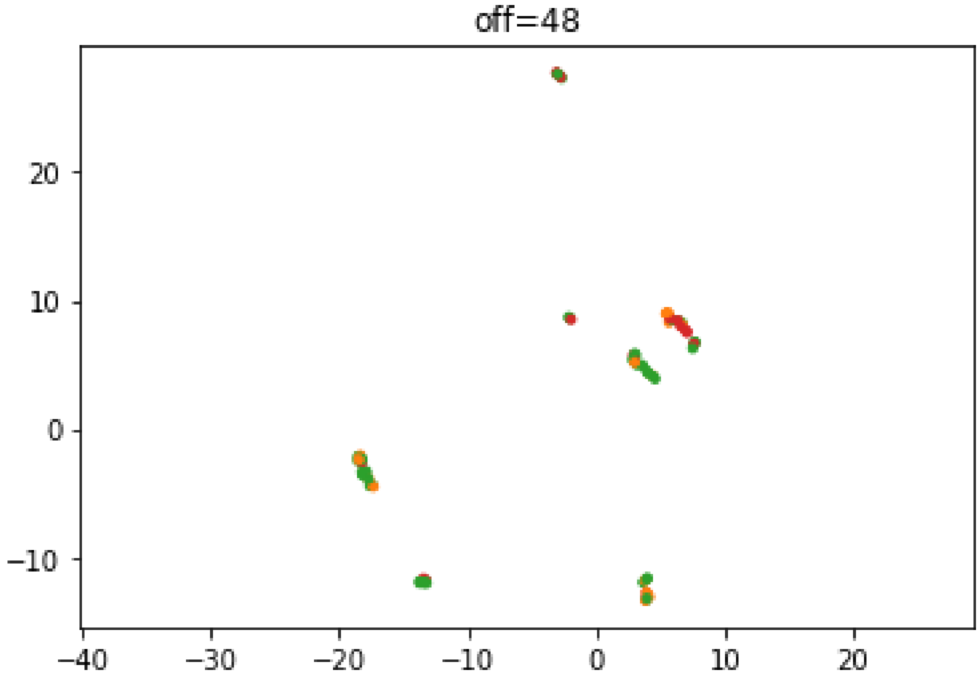
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 3 separate connected components using meta-embedding (experimental)
```



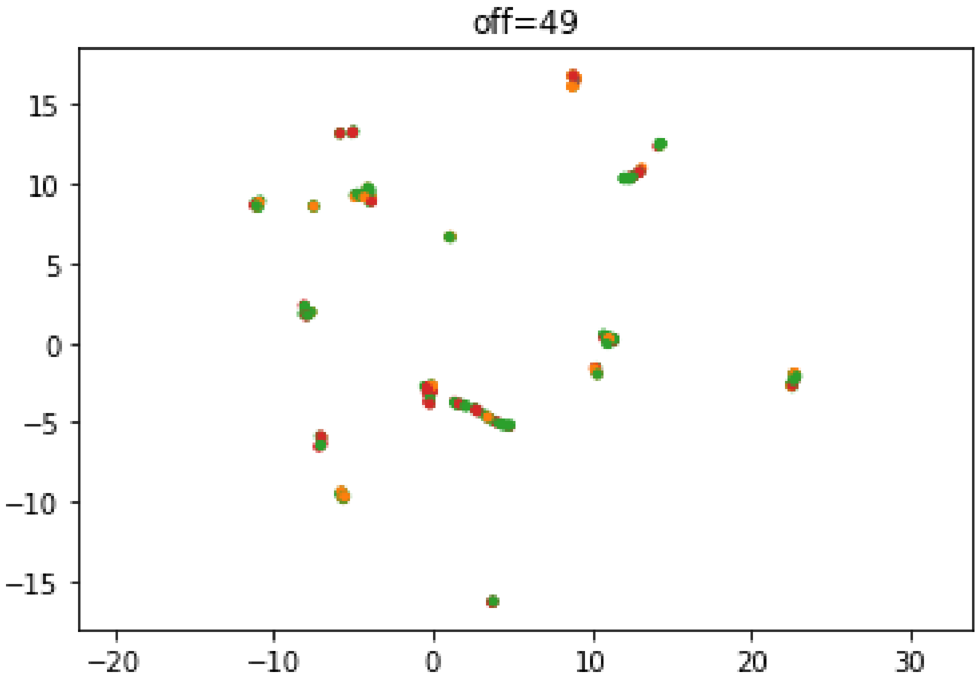
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

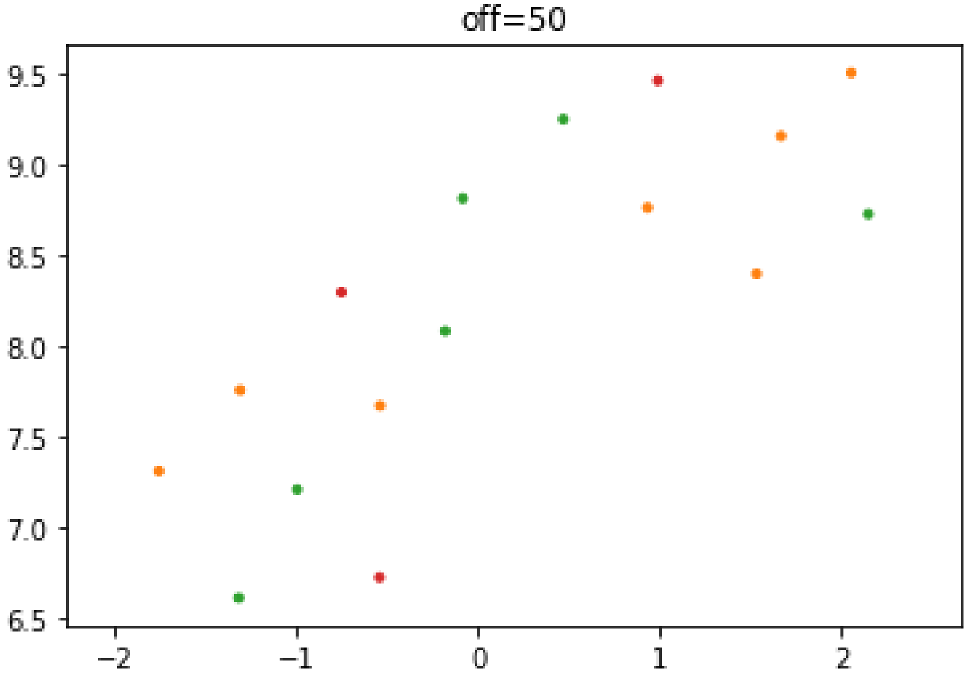


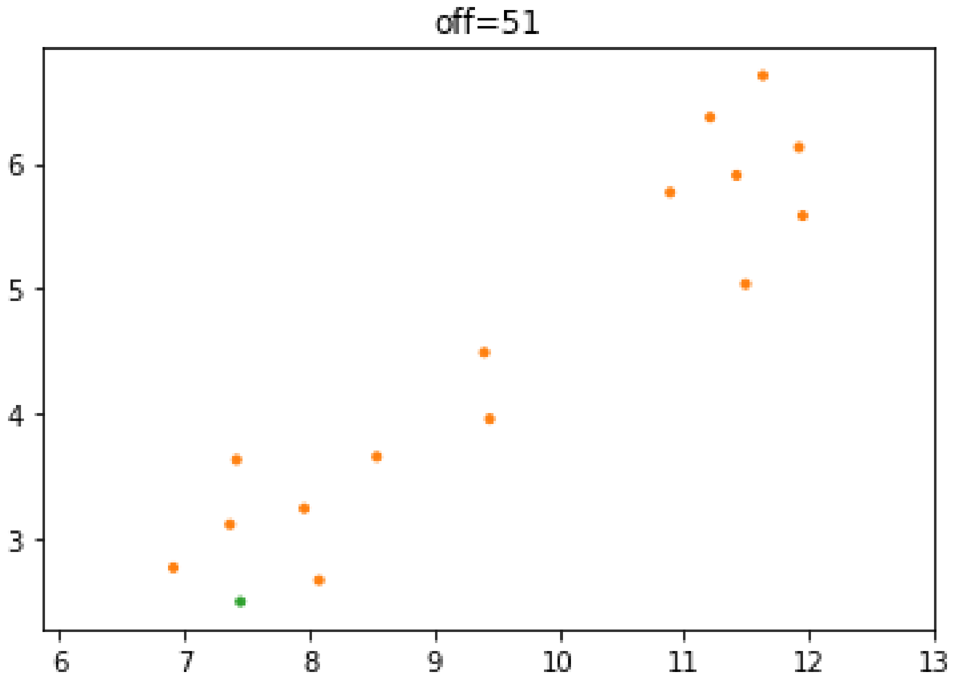
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 3 separate connected components using meta-embedding (experimental)
```



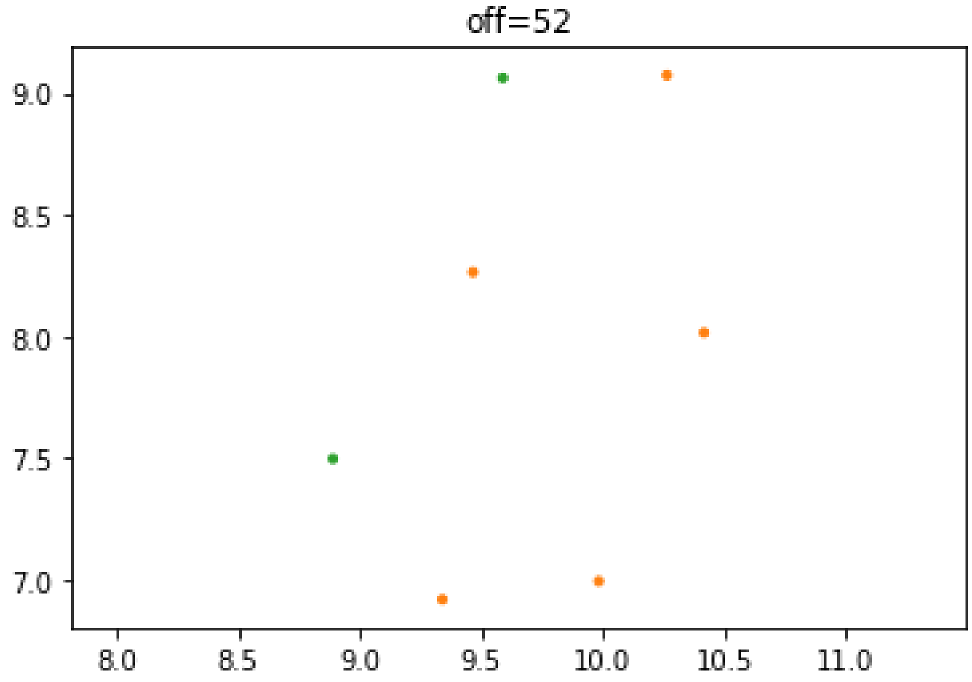
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 6 separate connected components using meta-embedding (experimental)

/home/ethan/.local/lib/python3.7/site-packages/sklearn/manifold/spectral_embedding_.py:235: Use

Graph is not fully connected, spectral embedding may not work as expected.
```

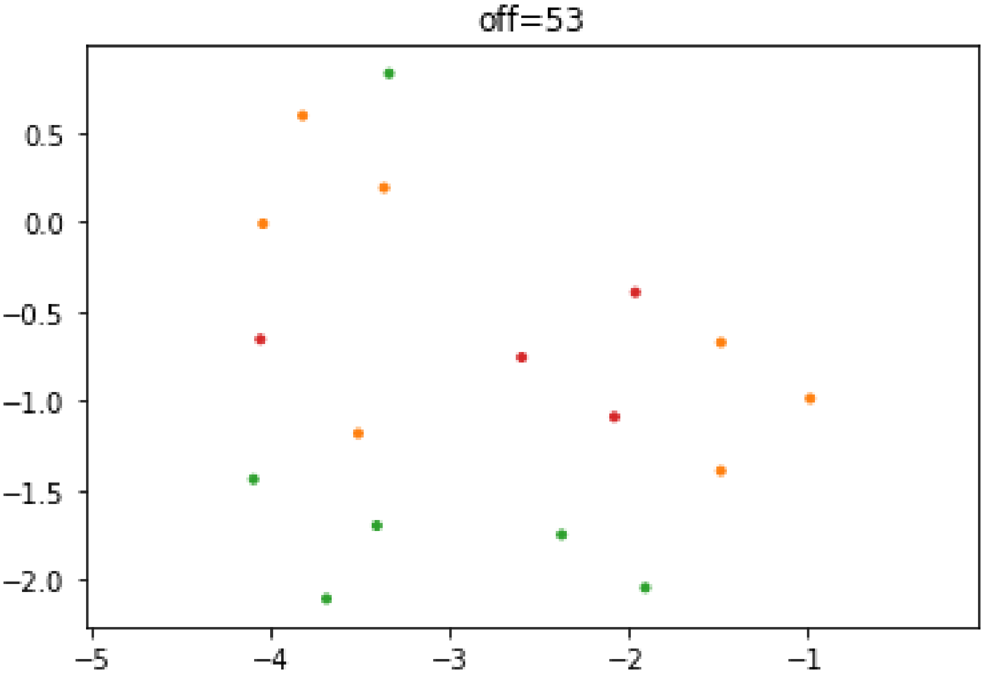

off=49
15
10
5
0
−5
−10
−15
−20
−10
0
10
20
30


off=50
9.5
9.0
8.5
8.0
7.5
7.0
6.5
−2
−1
0
1
2

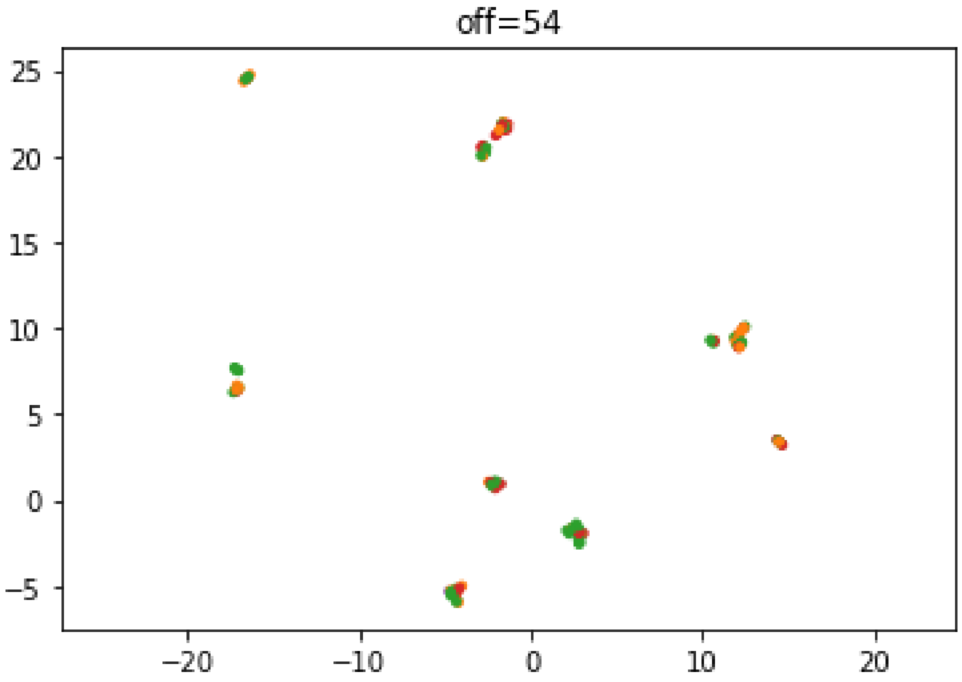


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

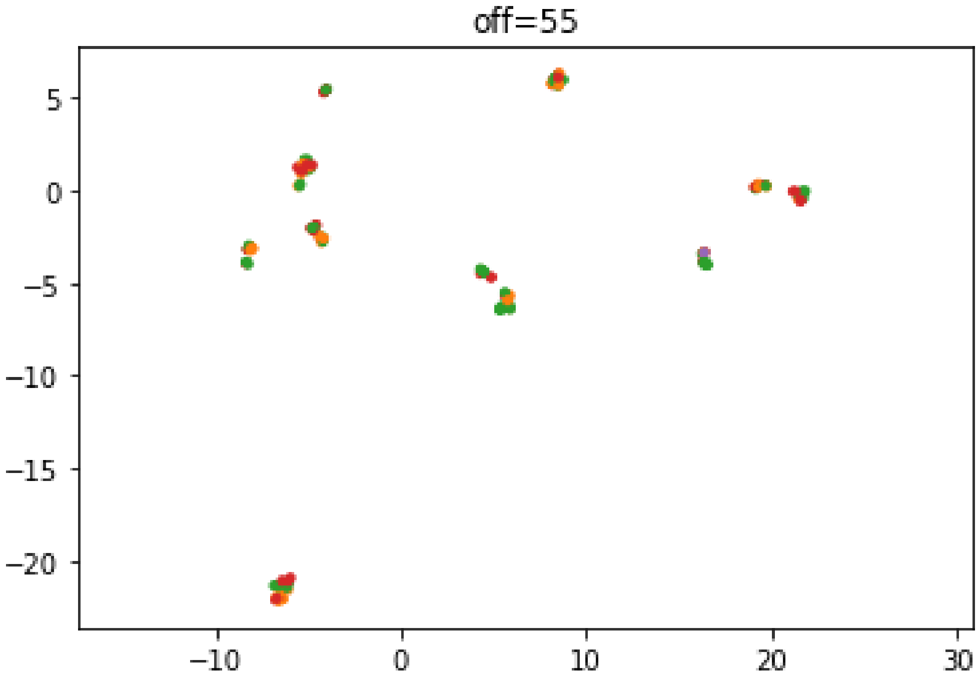


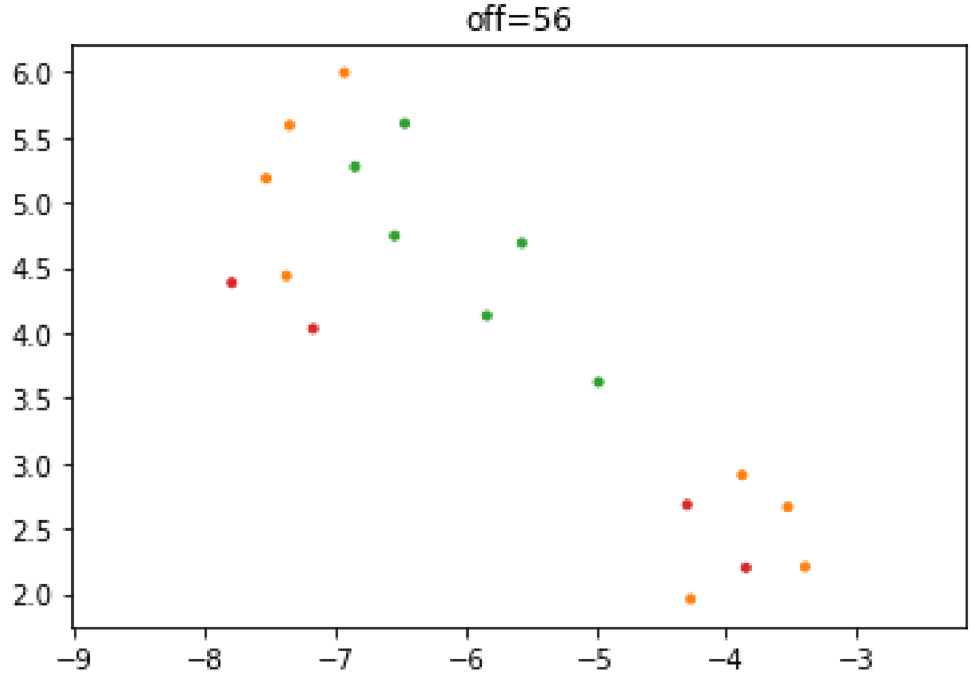


```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 3 separate connected components using meta-embedding (experimental)
```

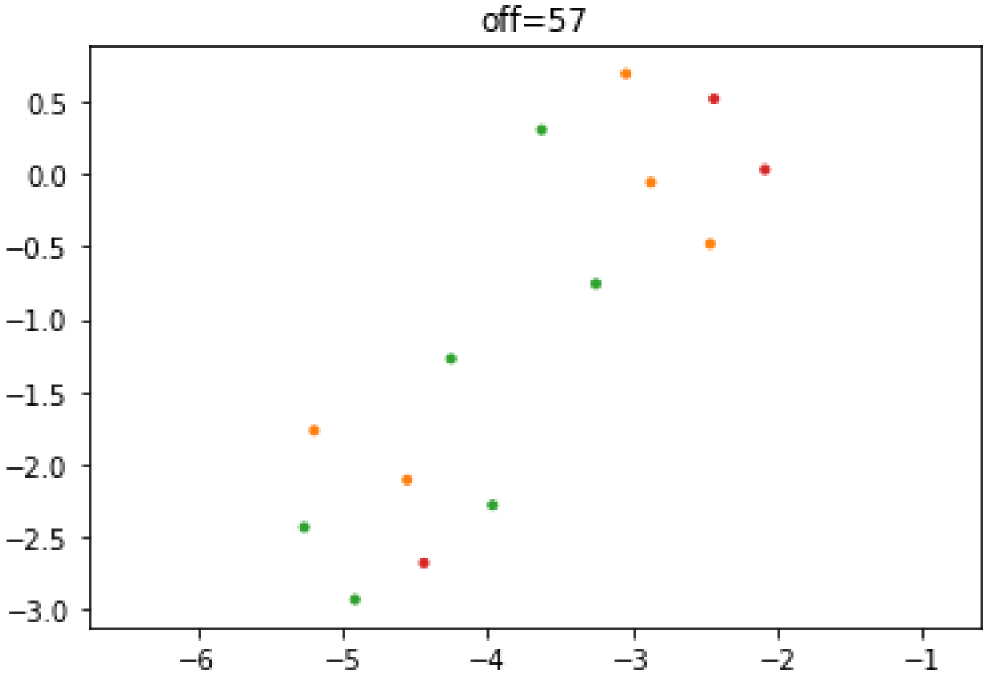


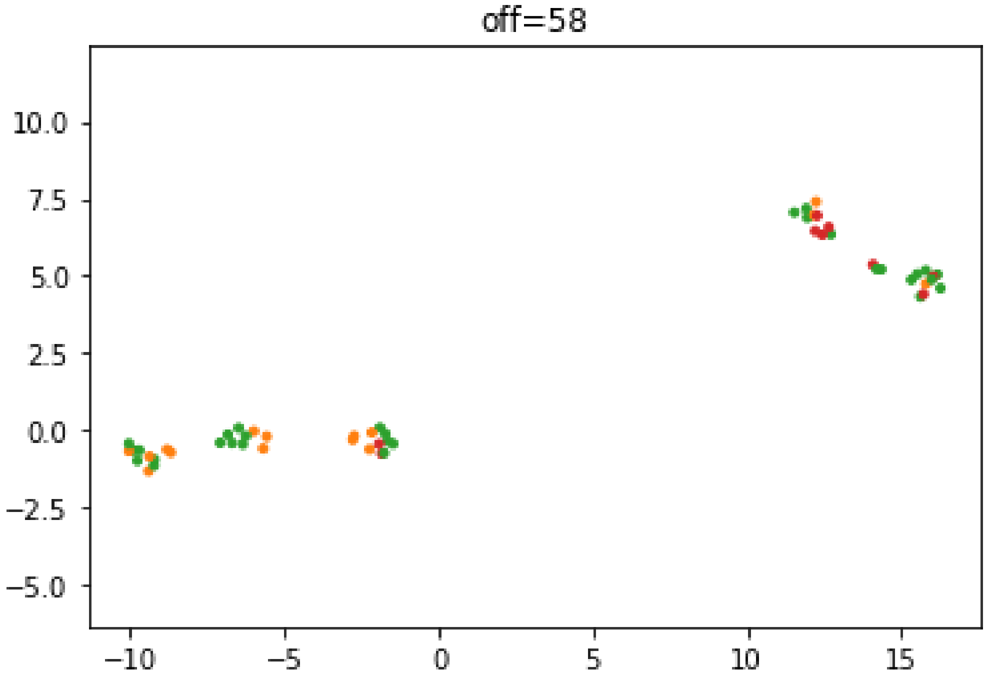
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 2 separate connected components using meta-embedding (experimental)
```





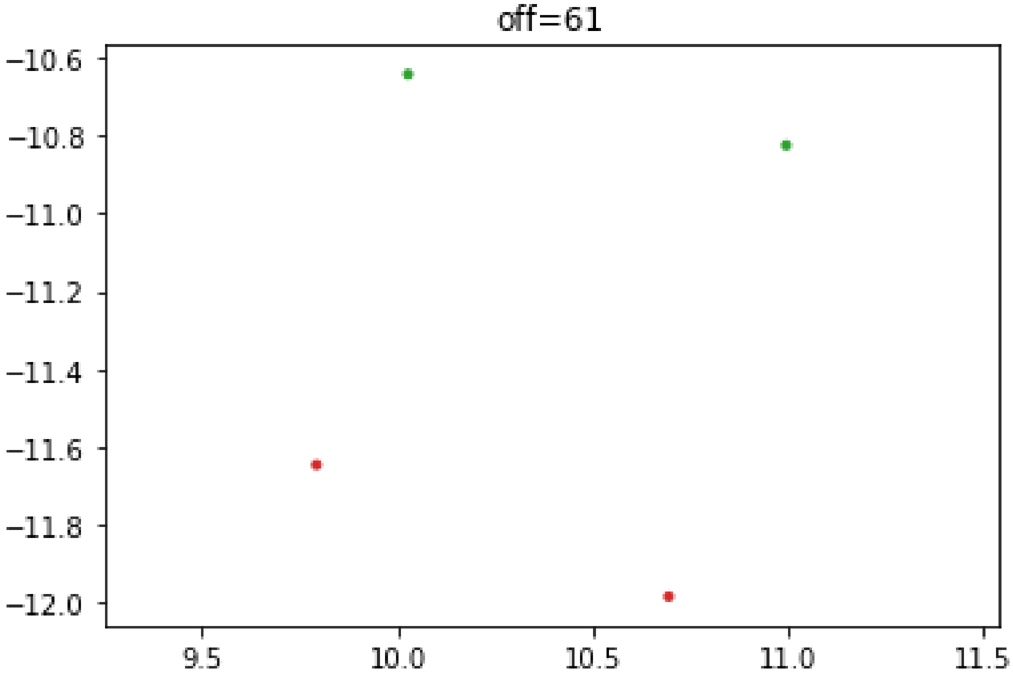
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

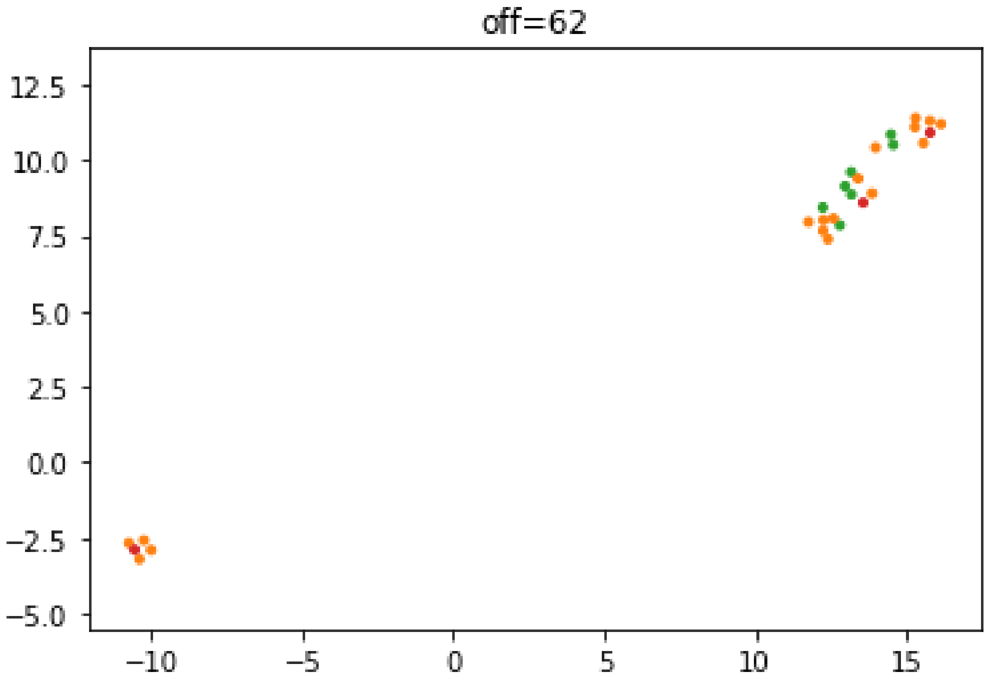


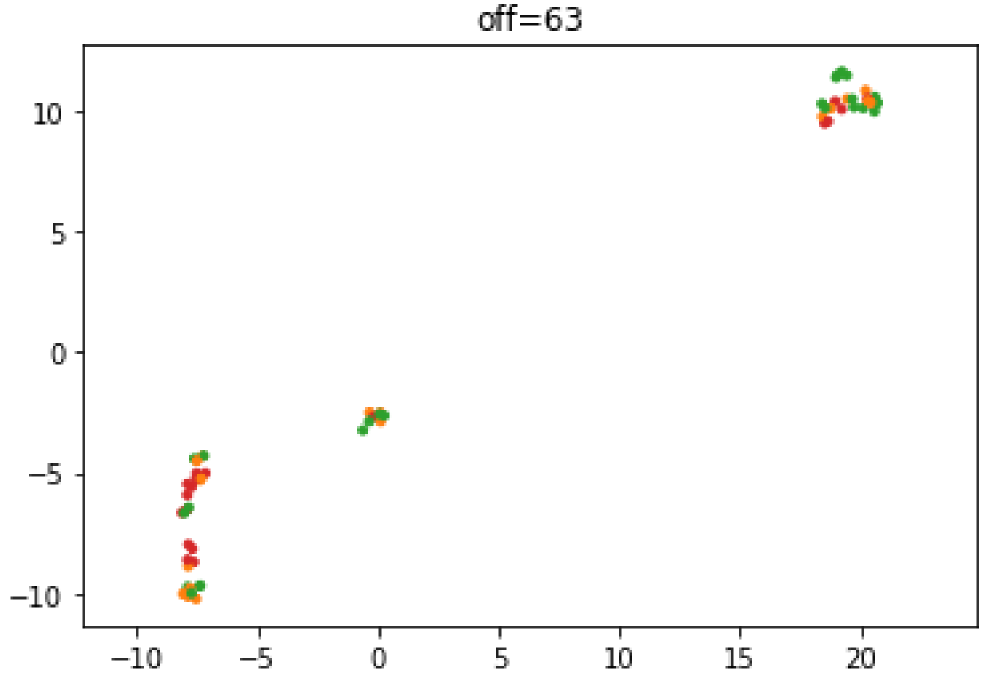


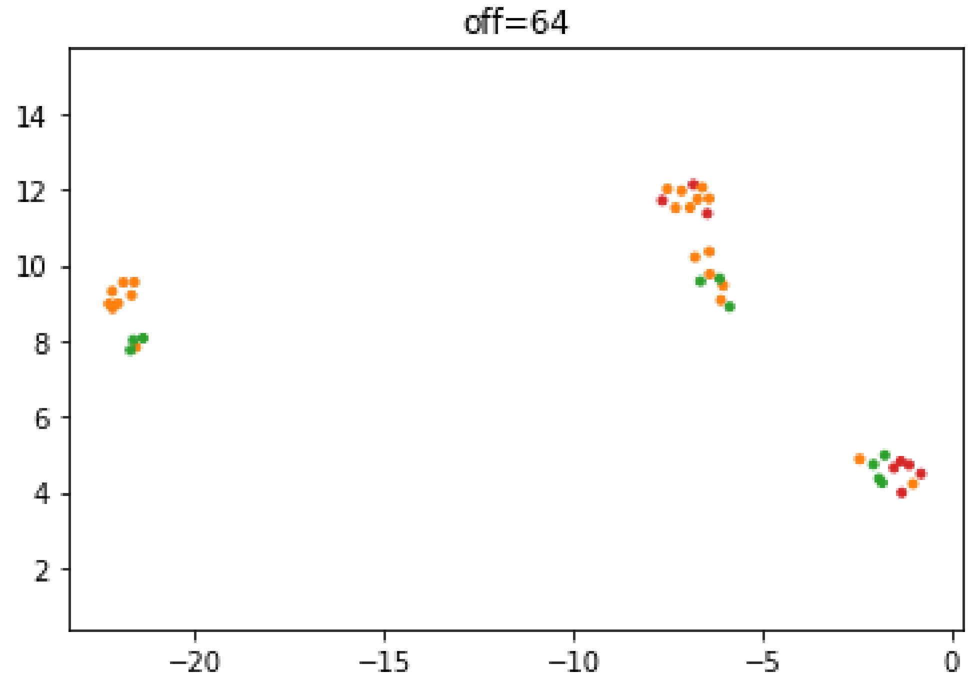
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

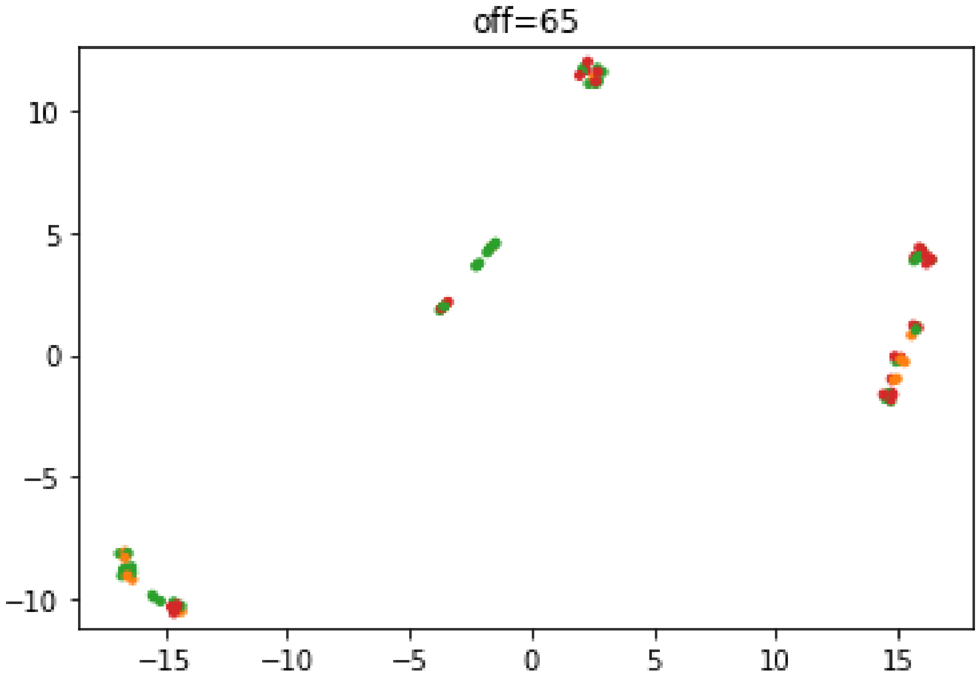



off=62
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−10
−5
0
5
10
15

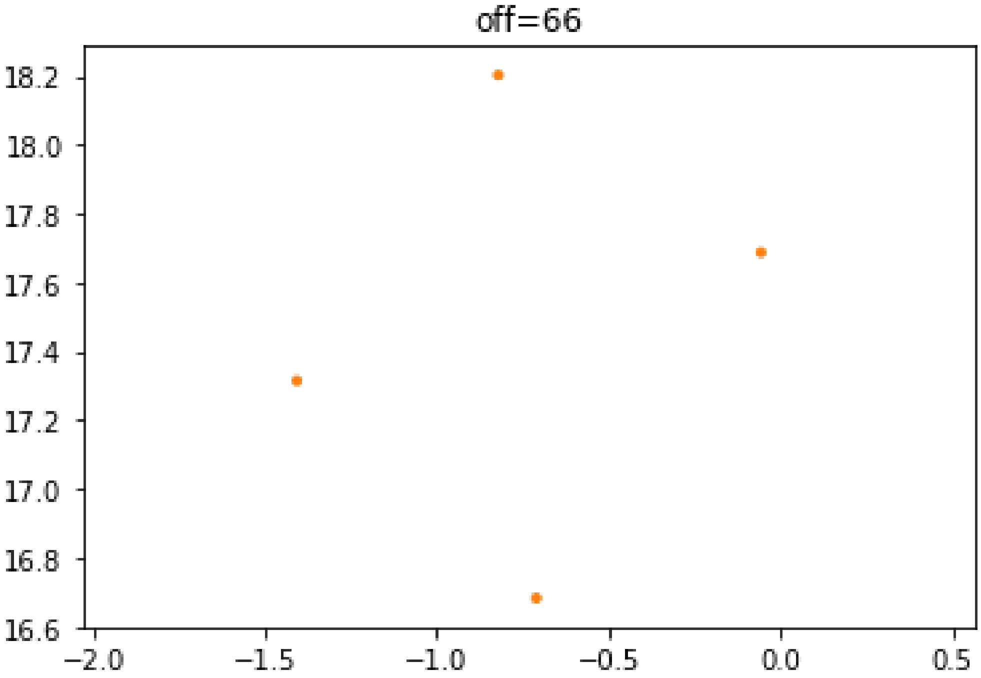

off=63
10
5
0
−5
−10
−10
−5
0
5
10
15
20



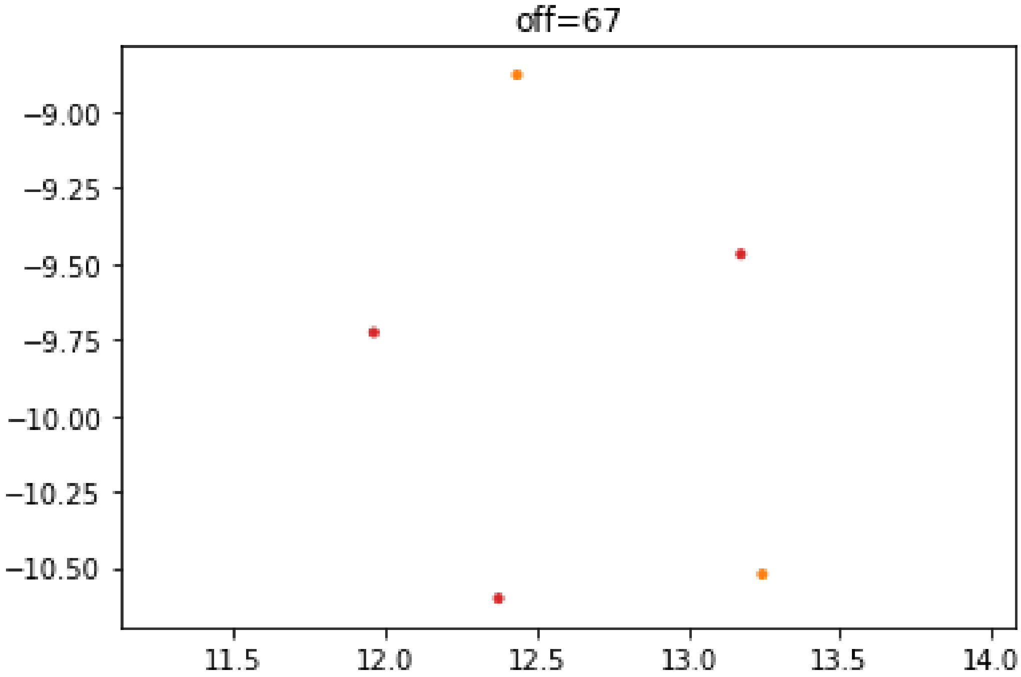
```
/home/ethan/.local/lib/python3.7/site-packages/umap/spectral.py:229: UserWarning:

Embedding a total of 3 separate connected components using meta-embedding (experimental)
```

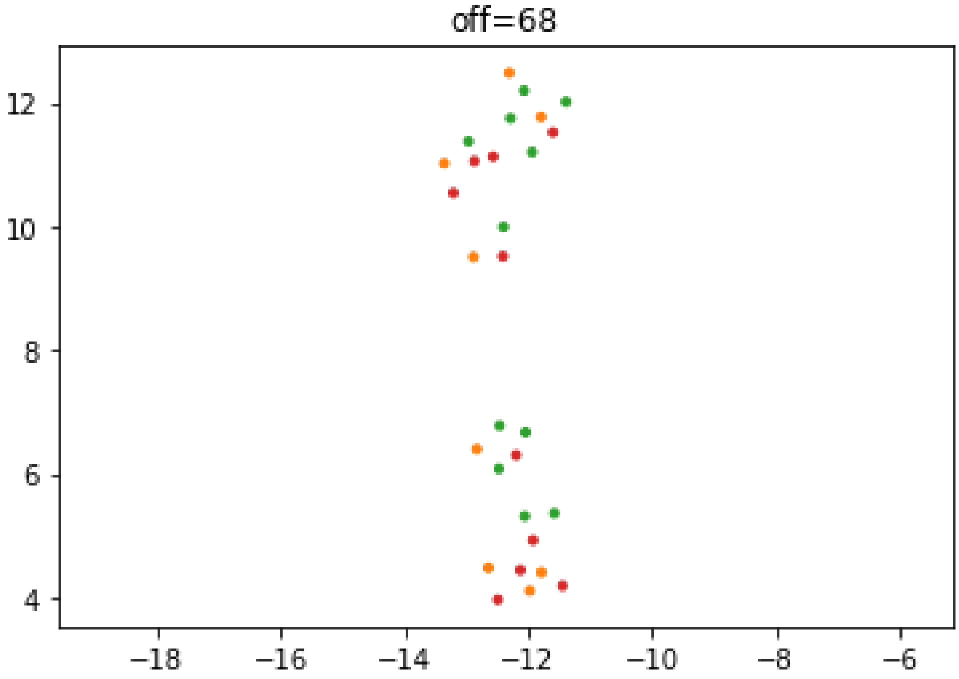


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



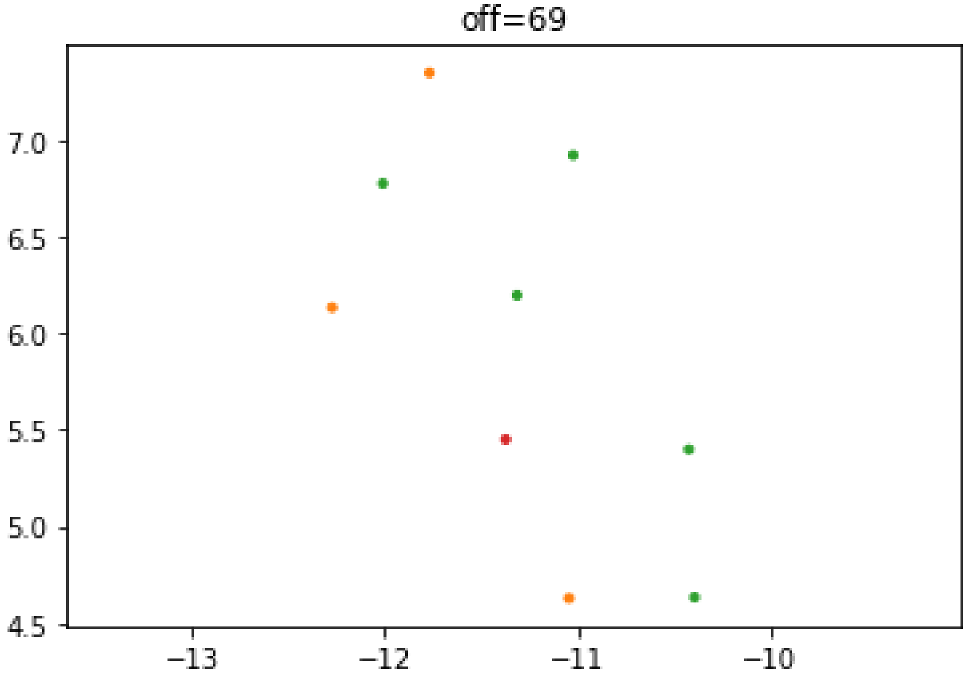
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



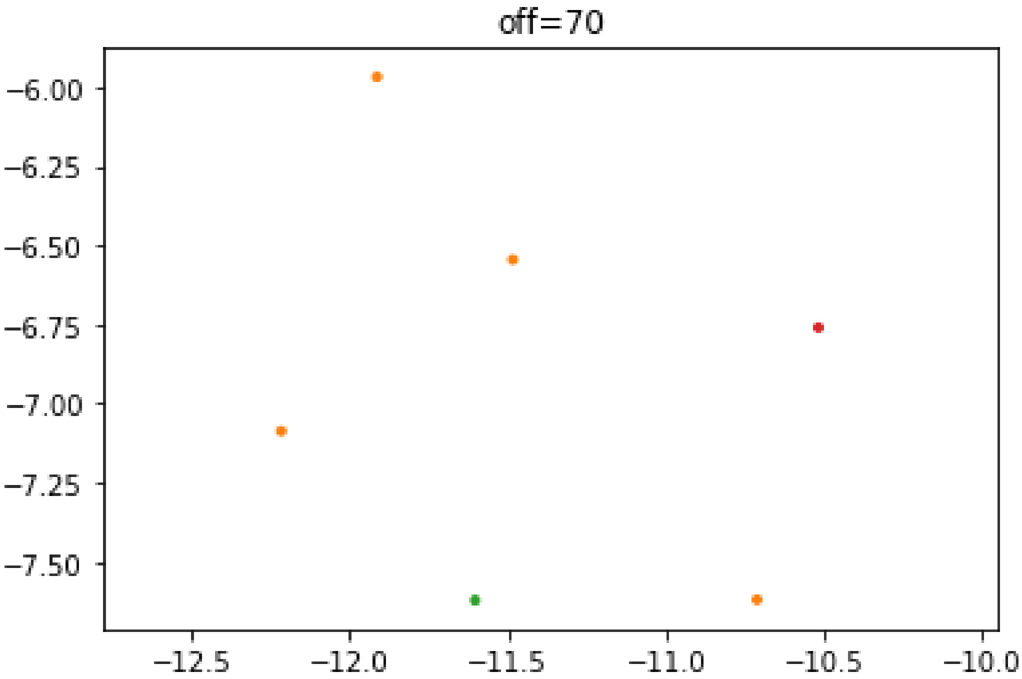


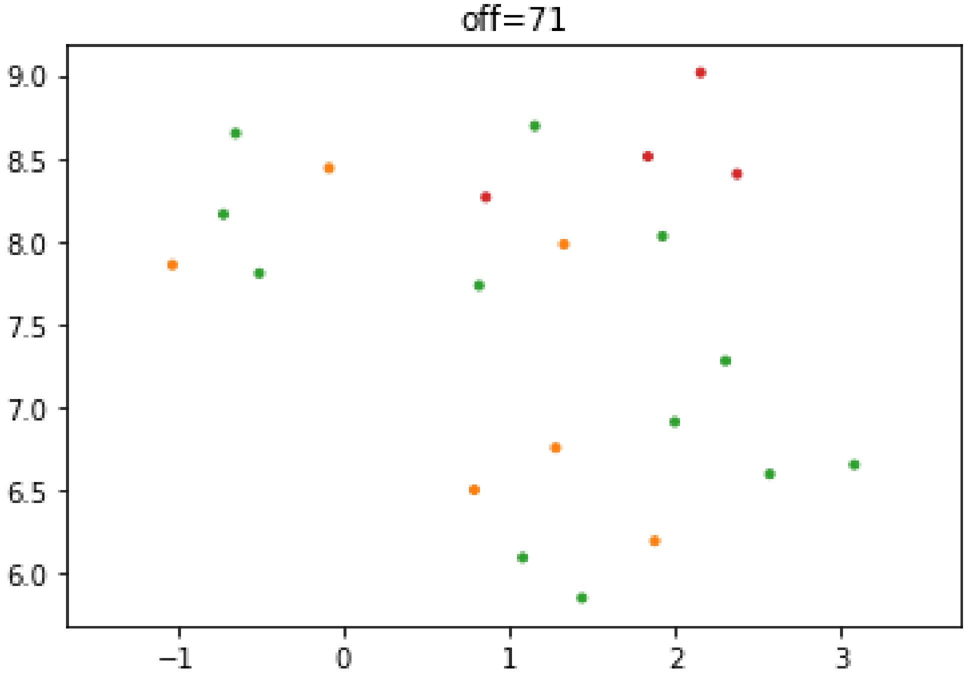
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

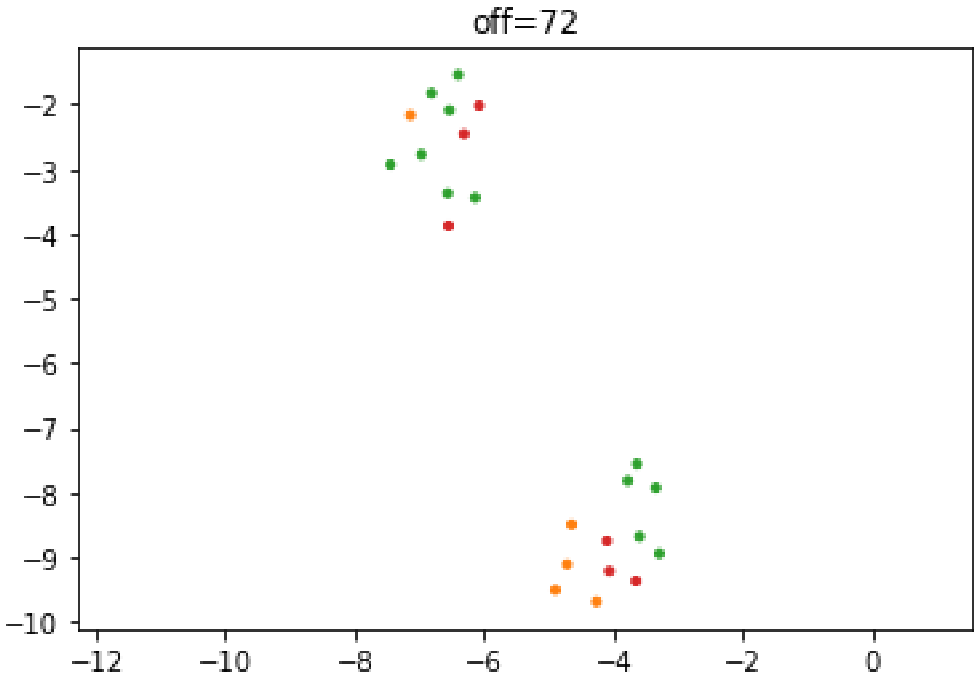
```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

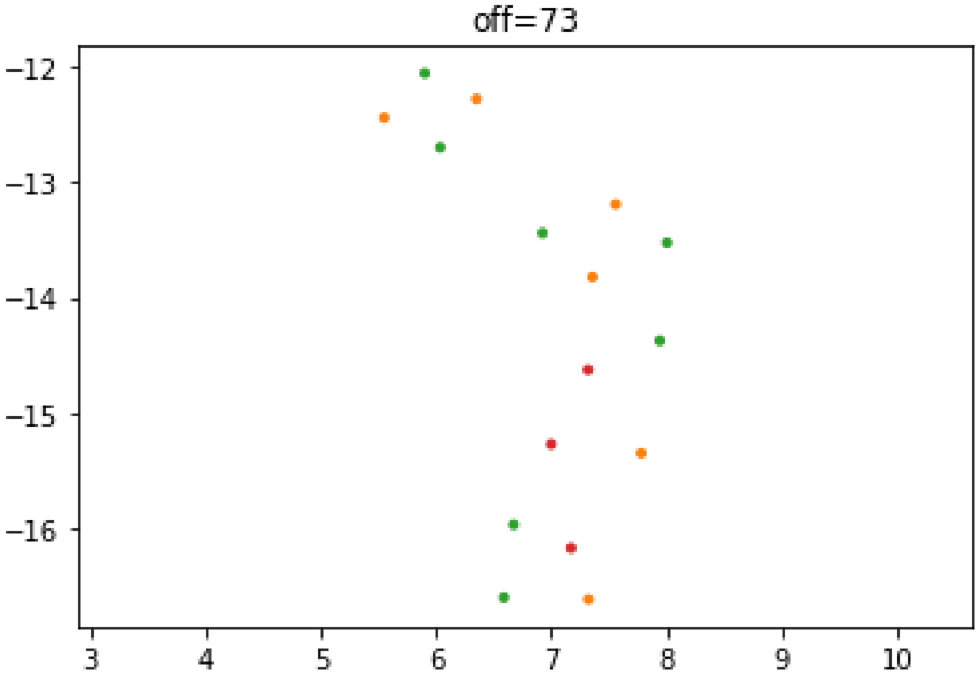


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```


off=70
−6.00
−6.25
−6.50
−6.75
−7.00
−7.25
−7.50
−12.5
−12.0
−11.5
−11.0
−10.5
−10.0

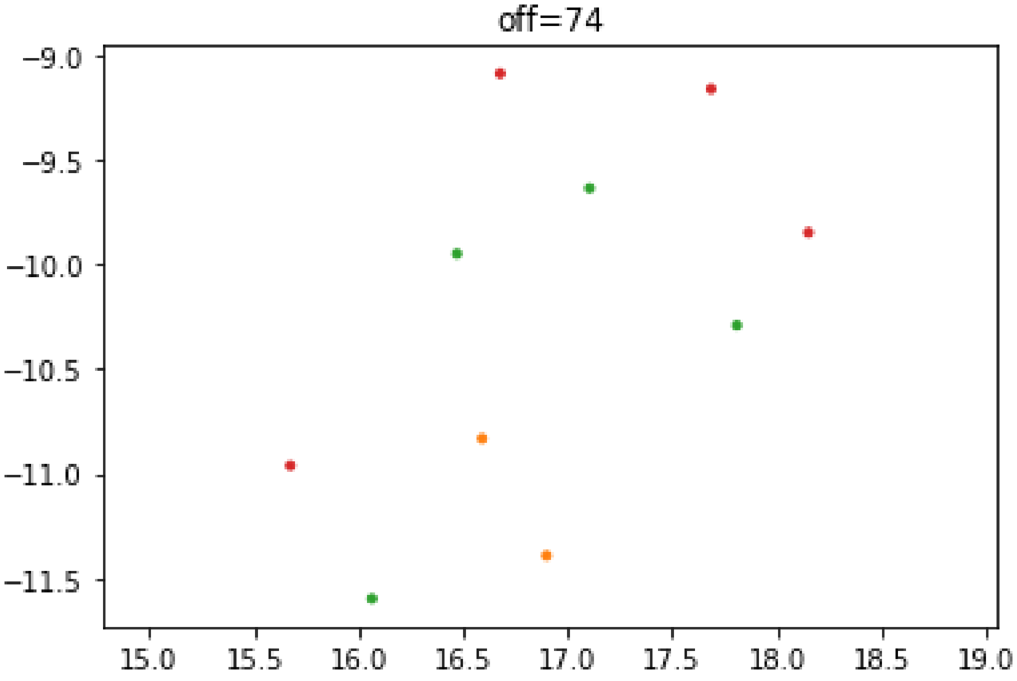

off=71
9.0
8.5
8.0
7.5
7.0
6.5
6.0
−1
0
1
2
3

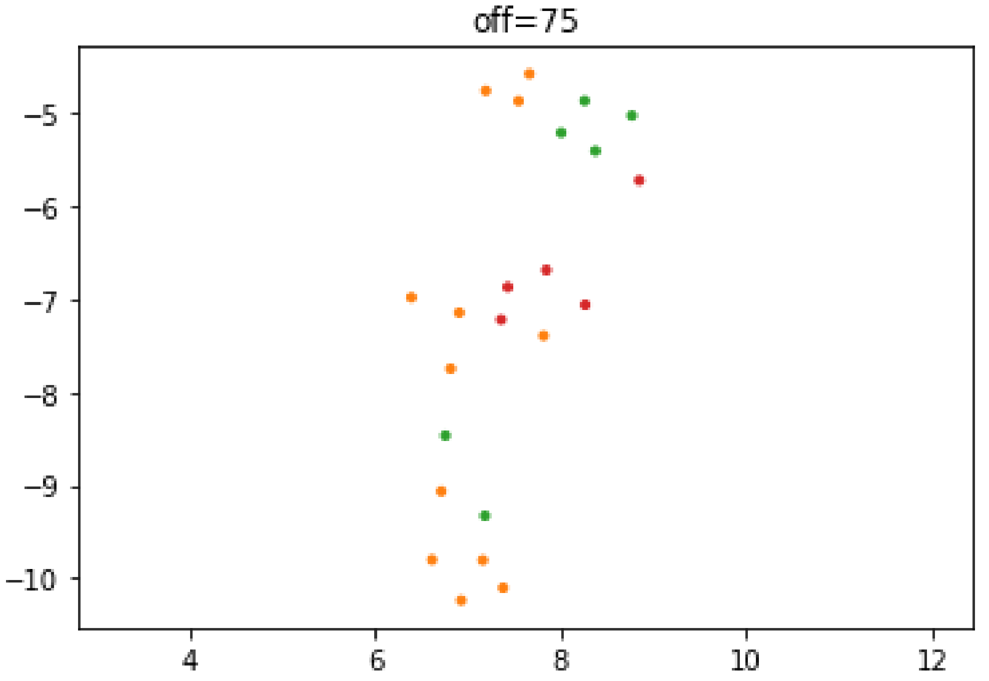




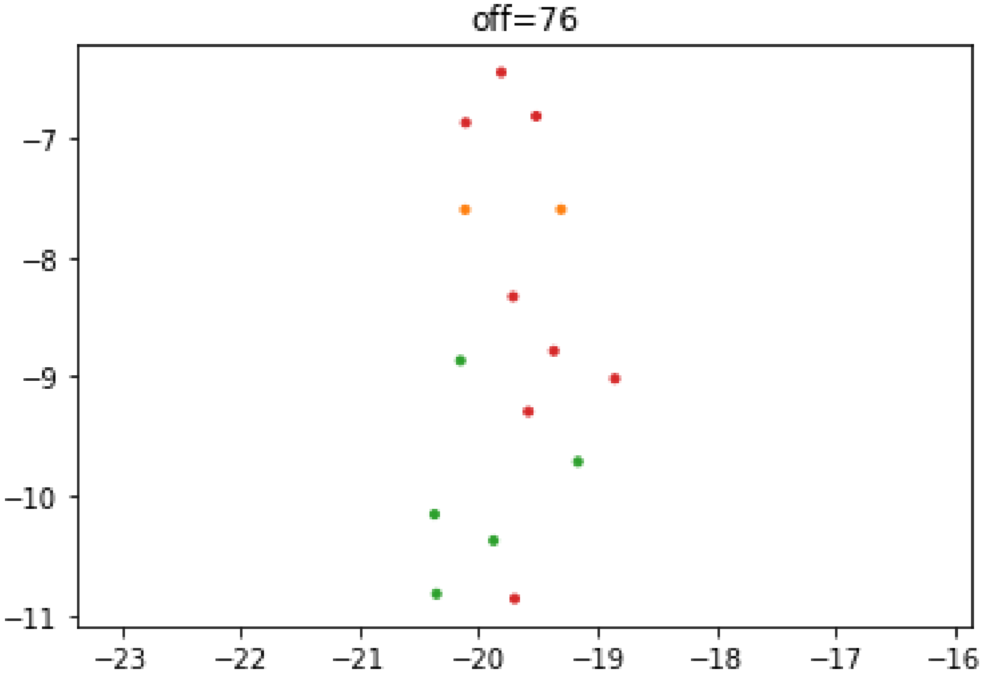
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

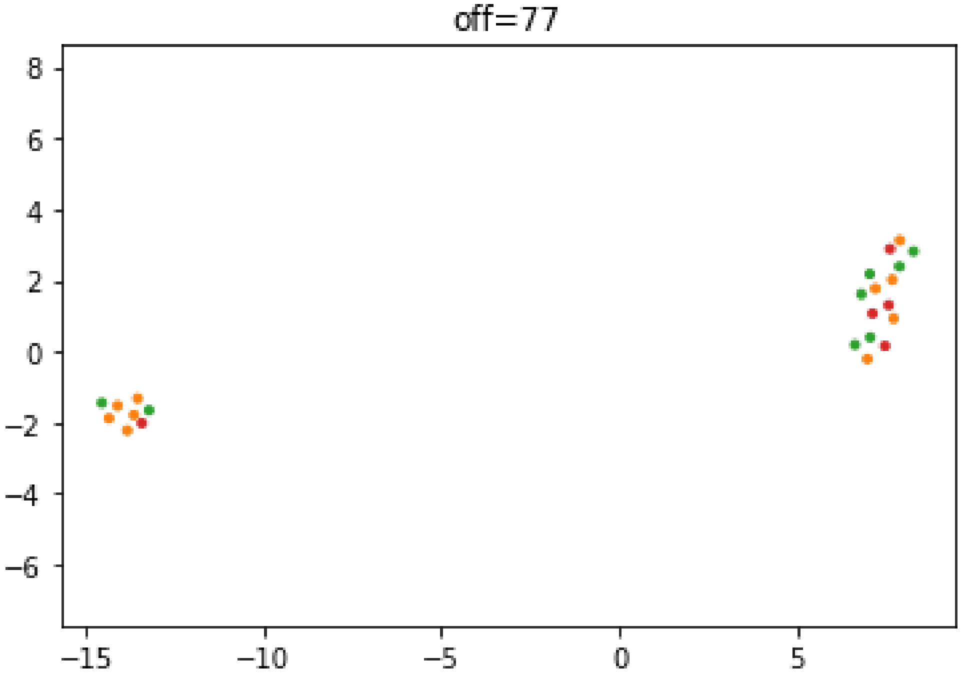
```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



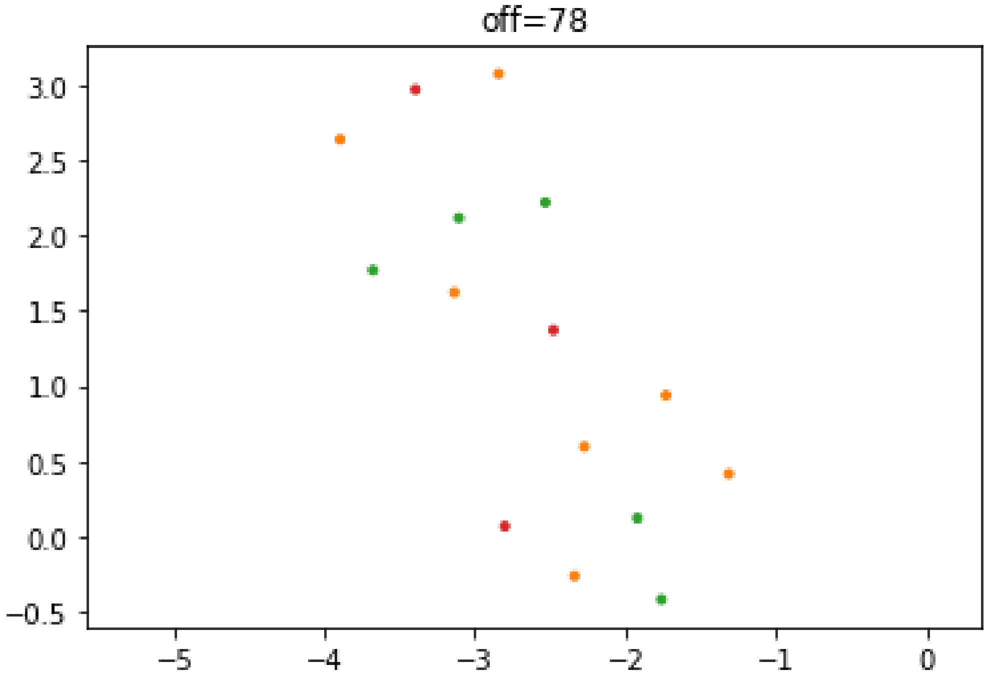


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

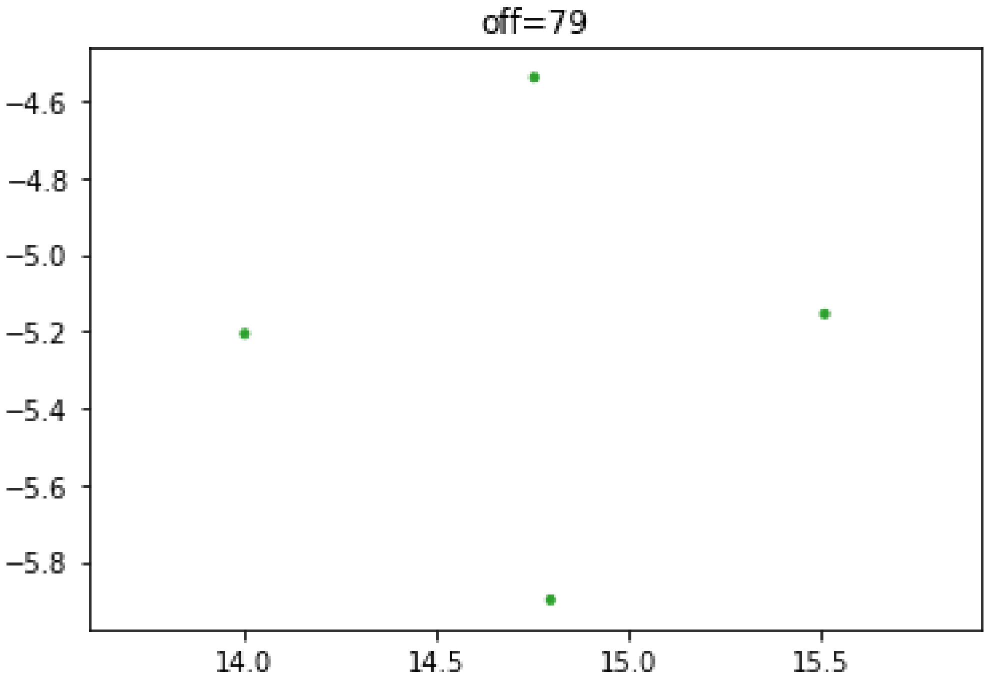




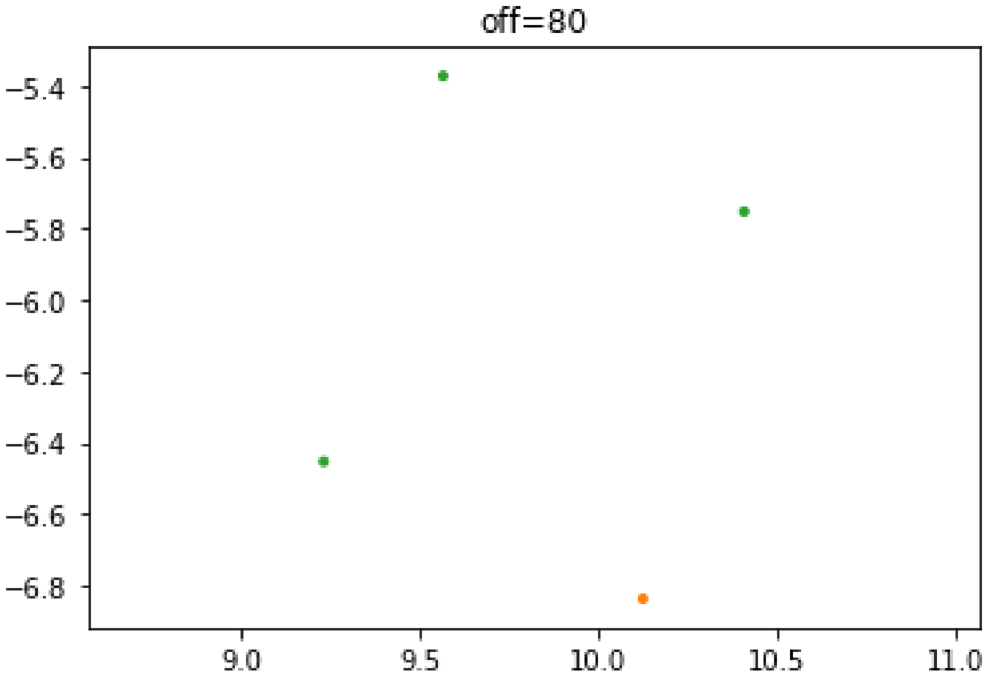
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



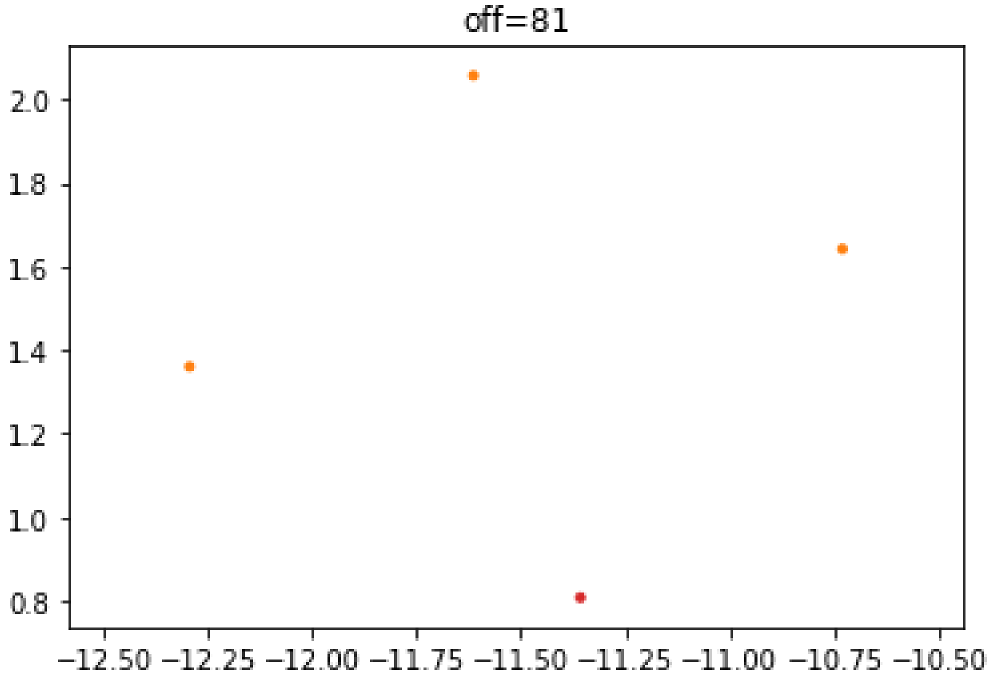
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

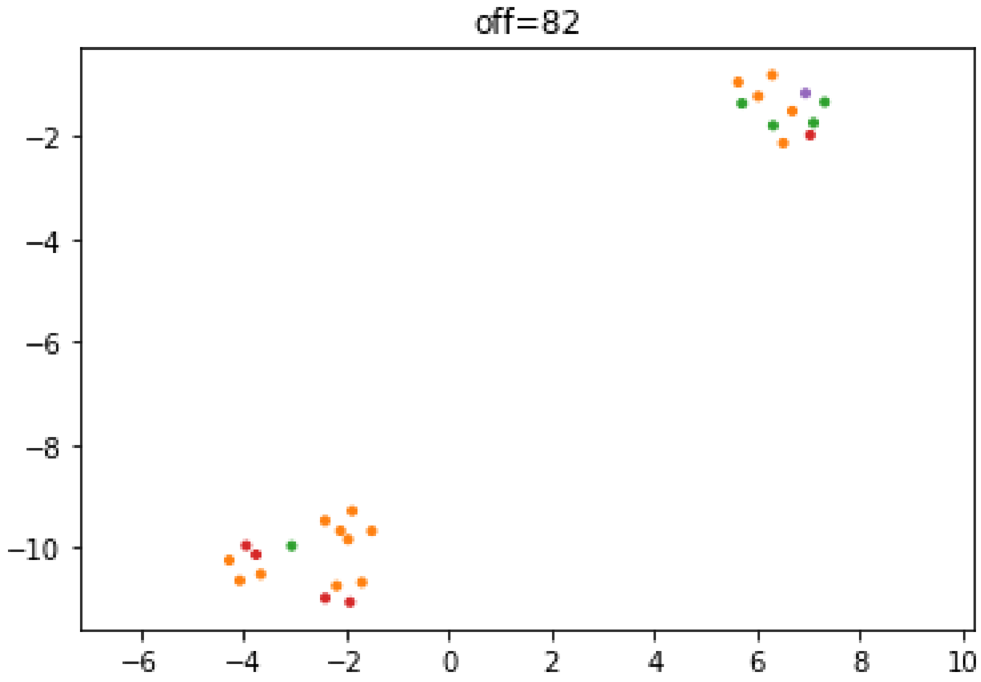


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

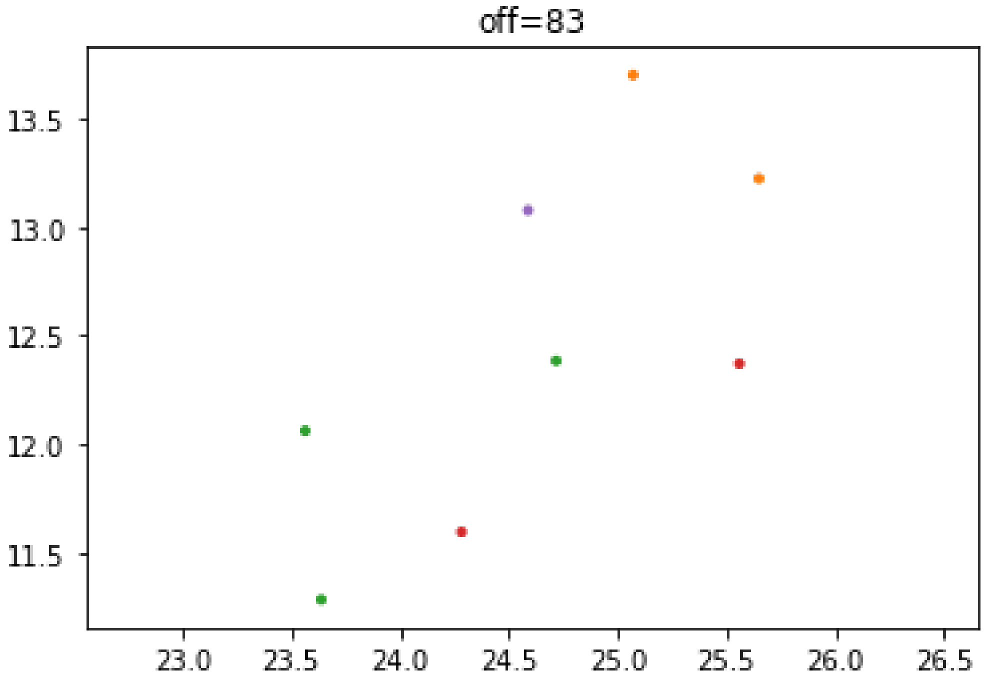


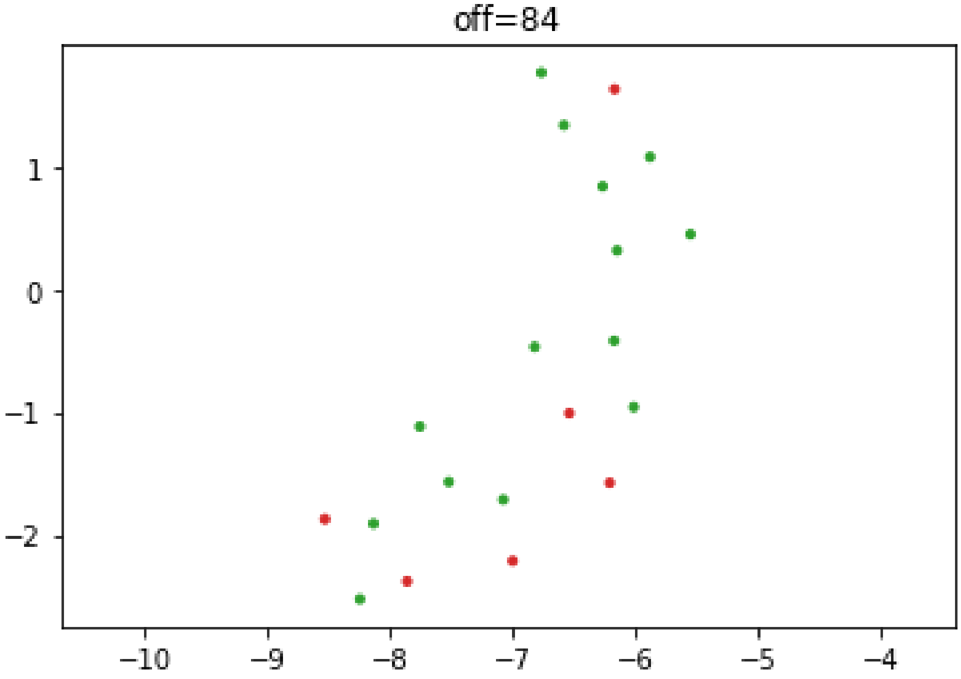
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



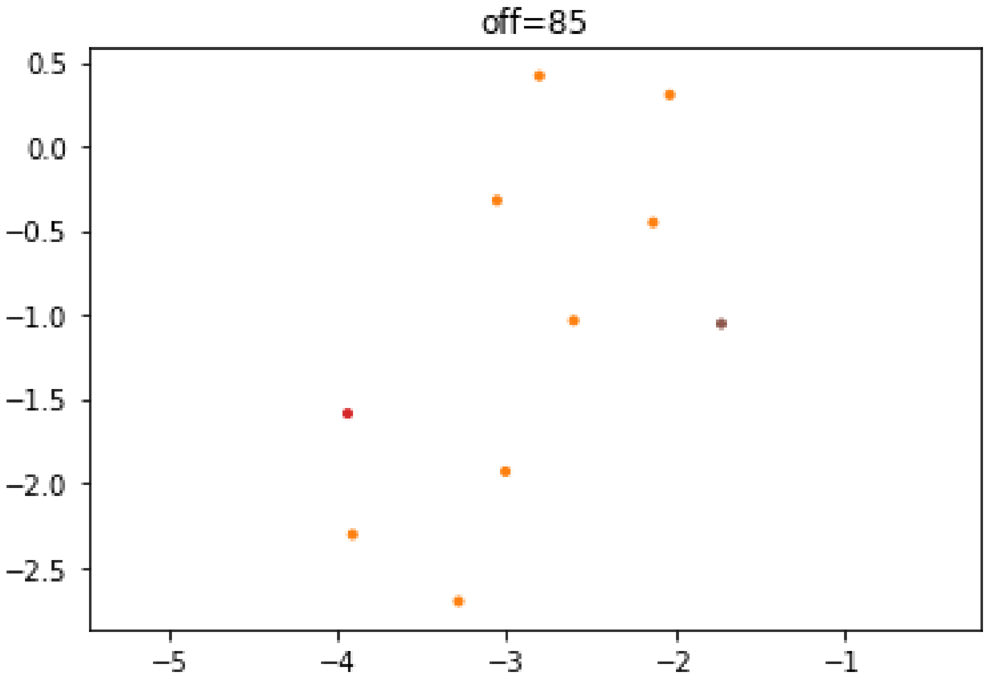


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

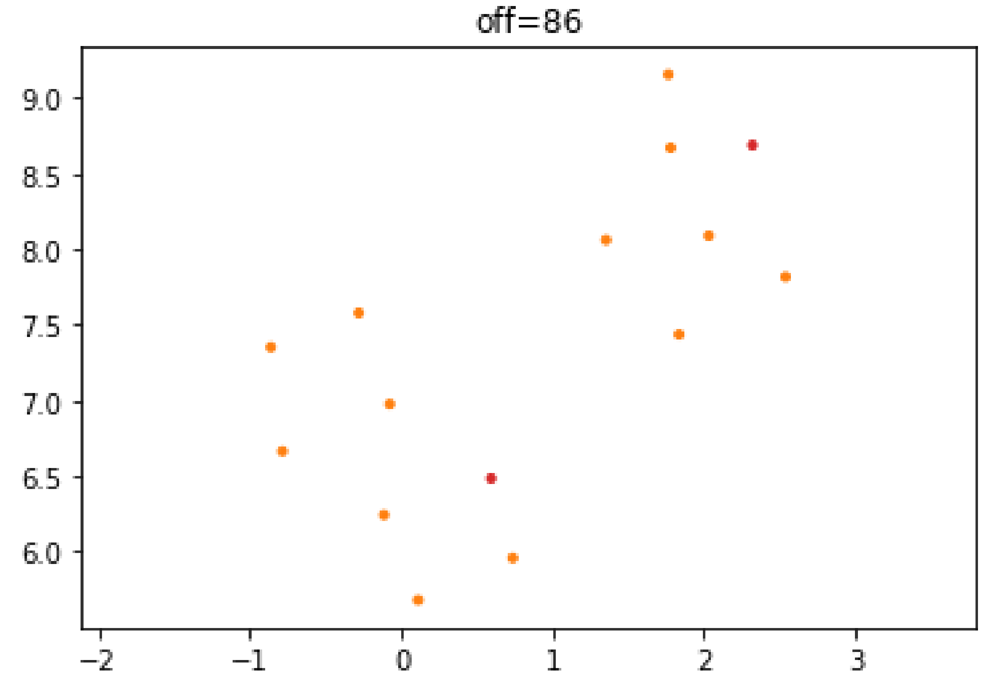




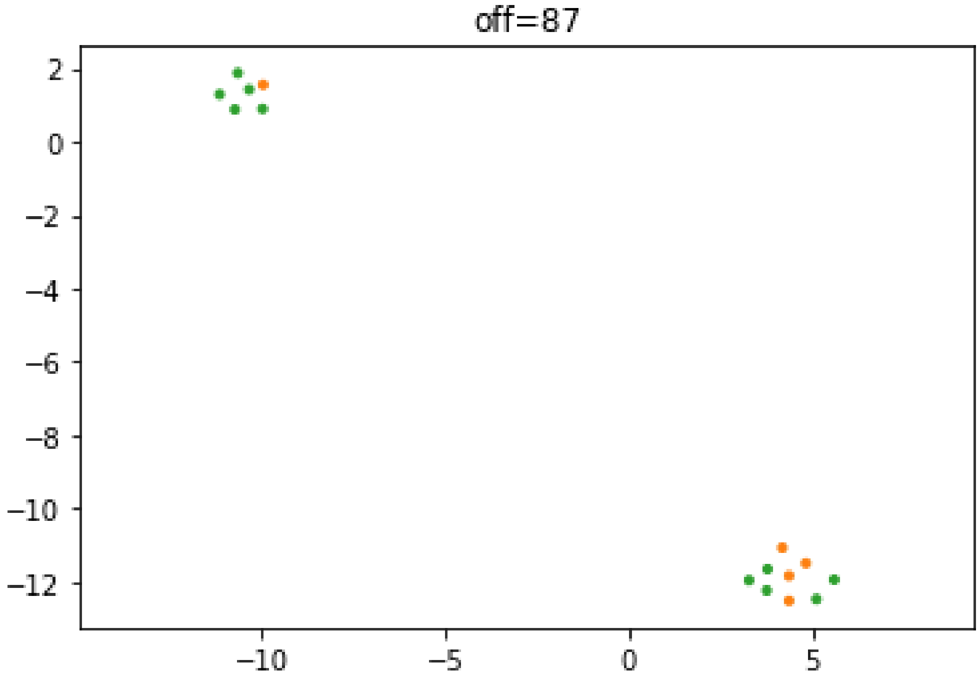
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



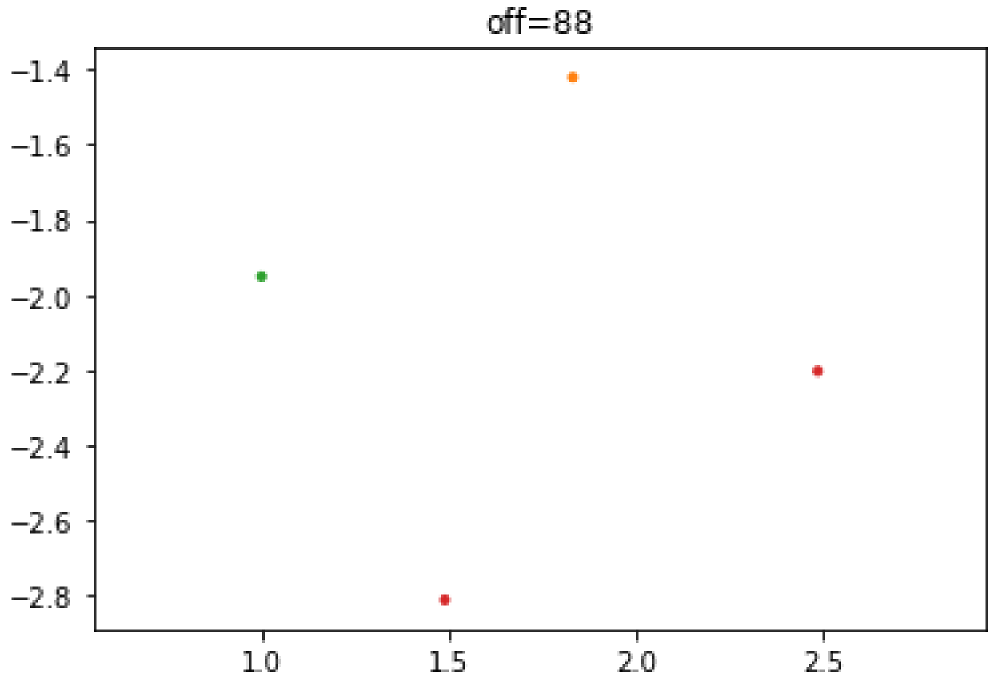
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



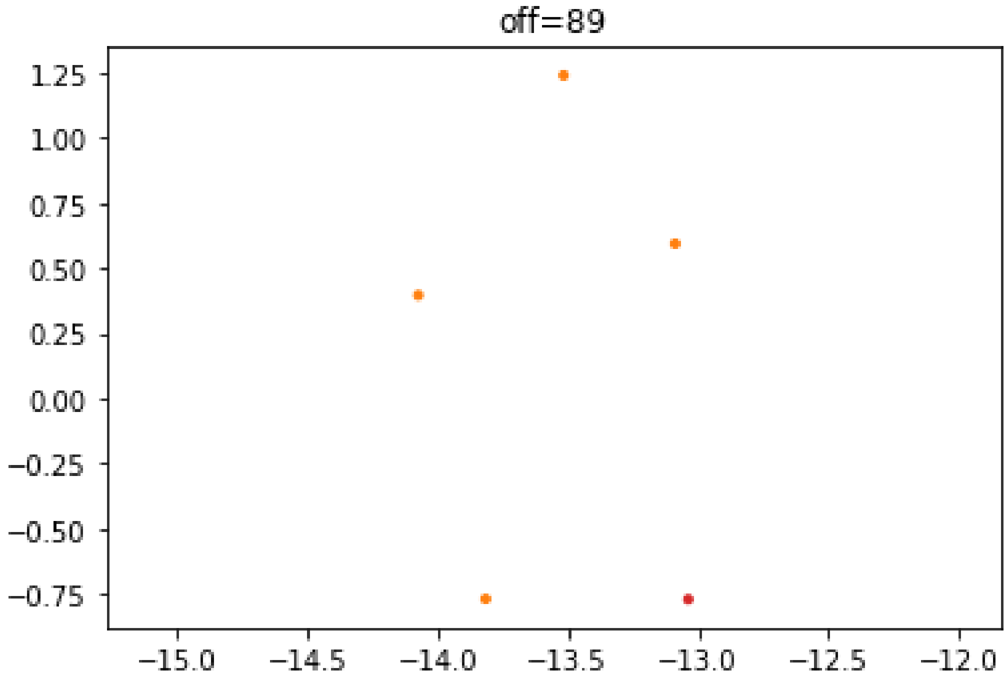
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

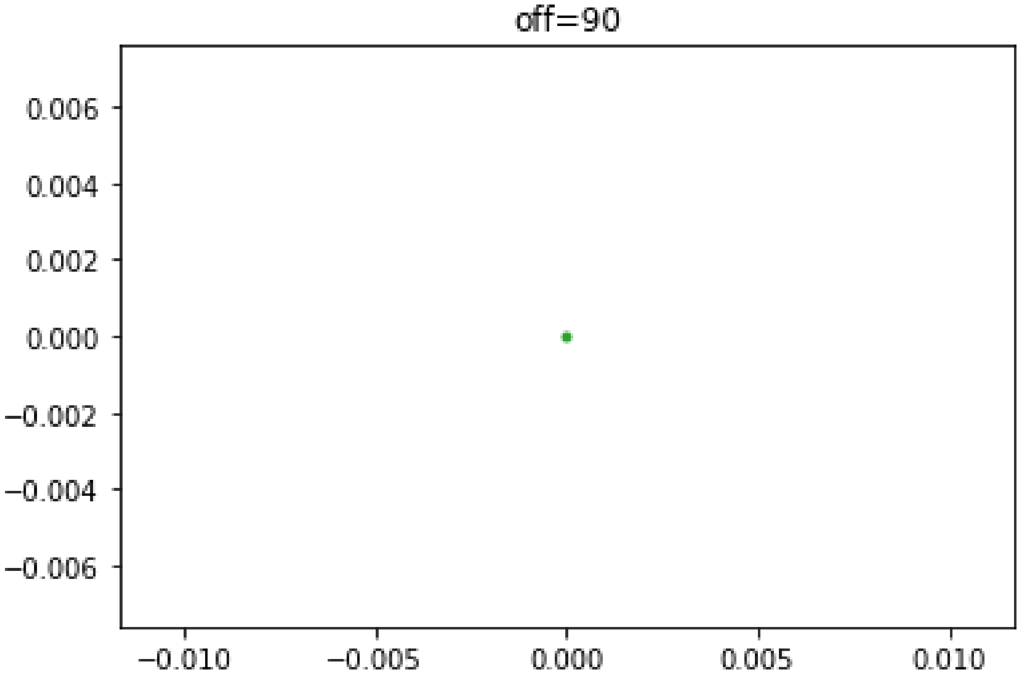


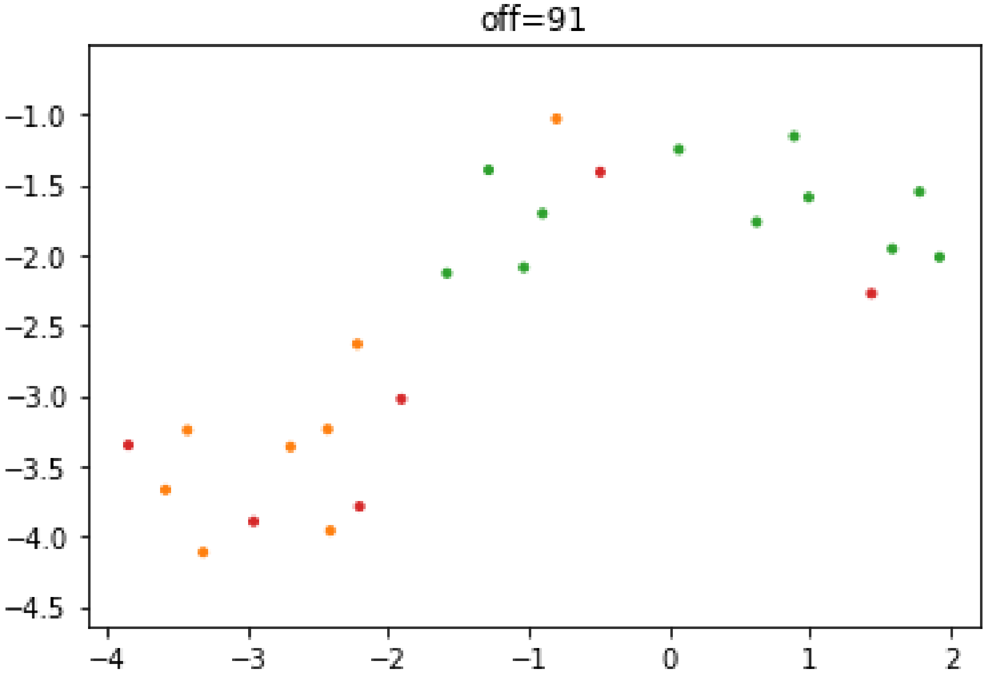
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



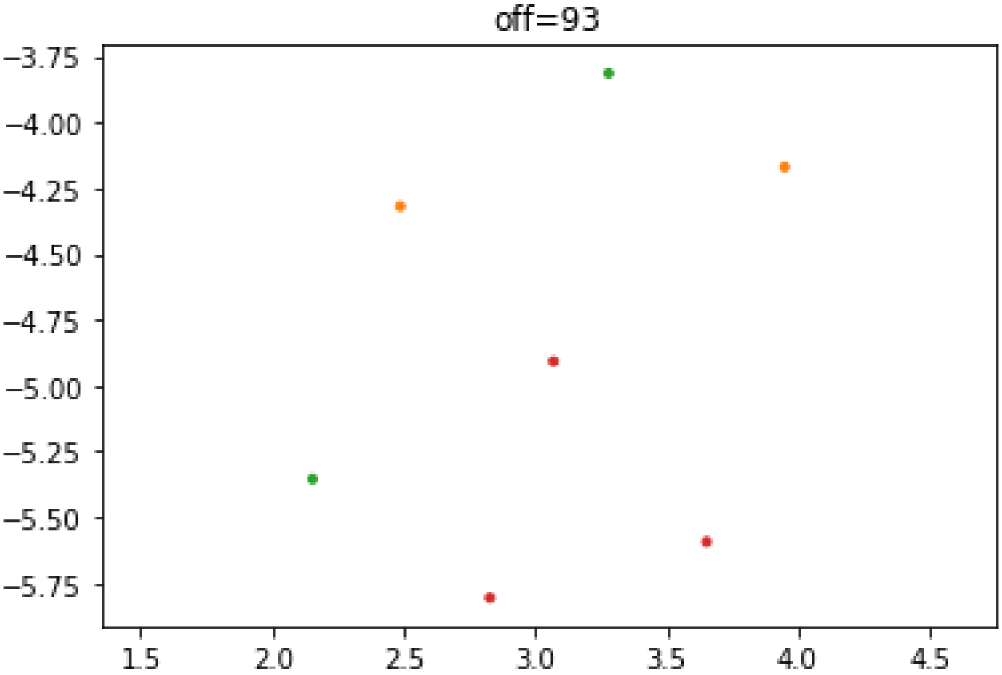
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



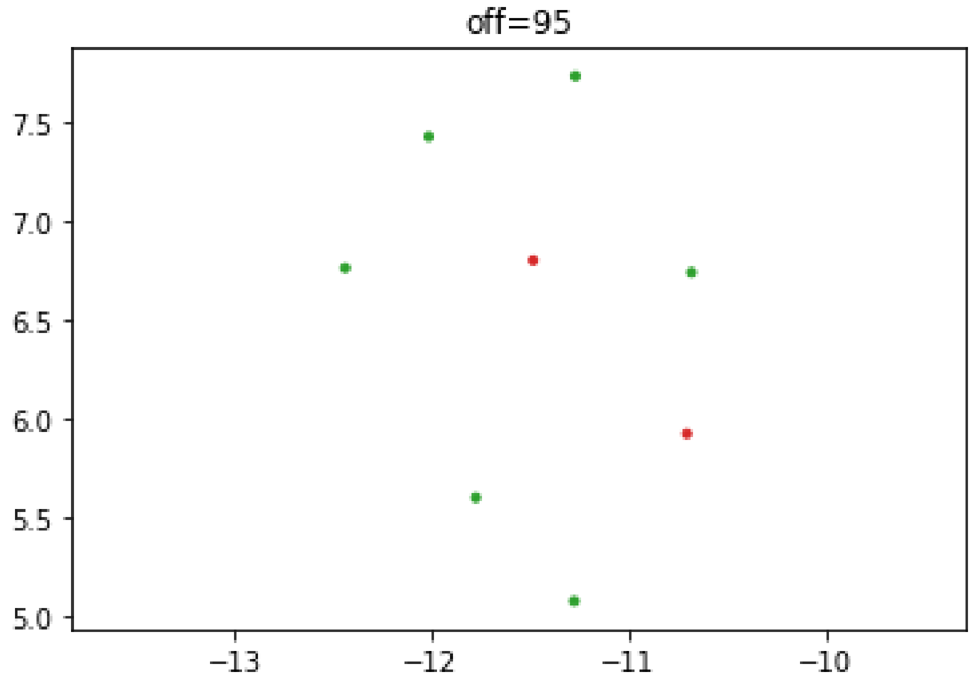

off=90
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010


off=91
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
−4.5
−4
−3
−2
−1
0
1
2

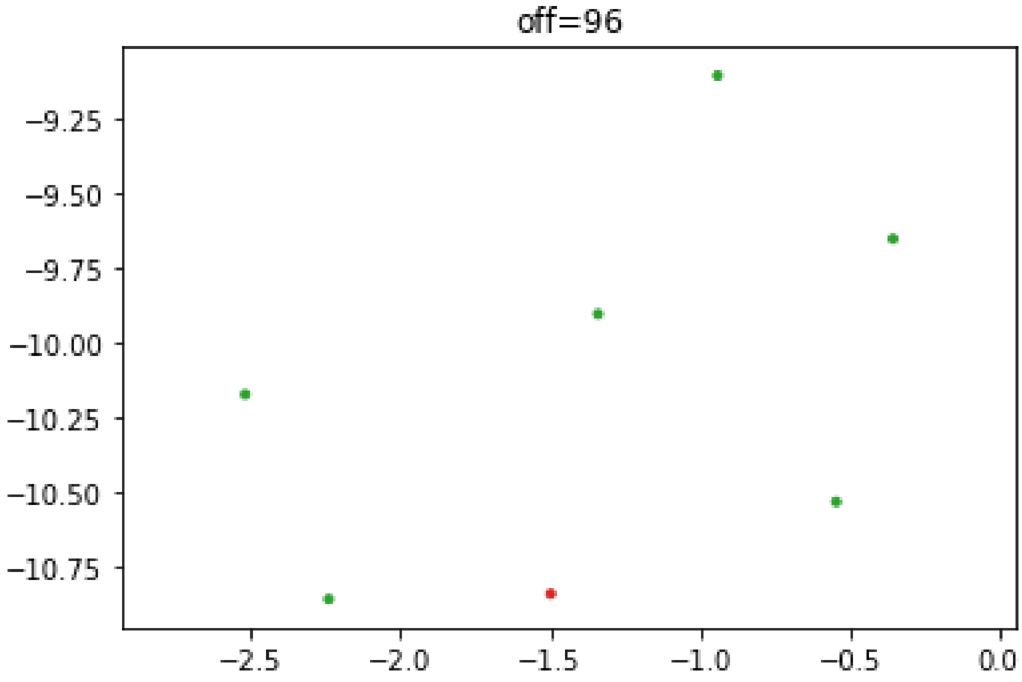
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



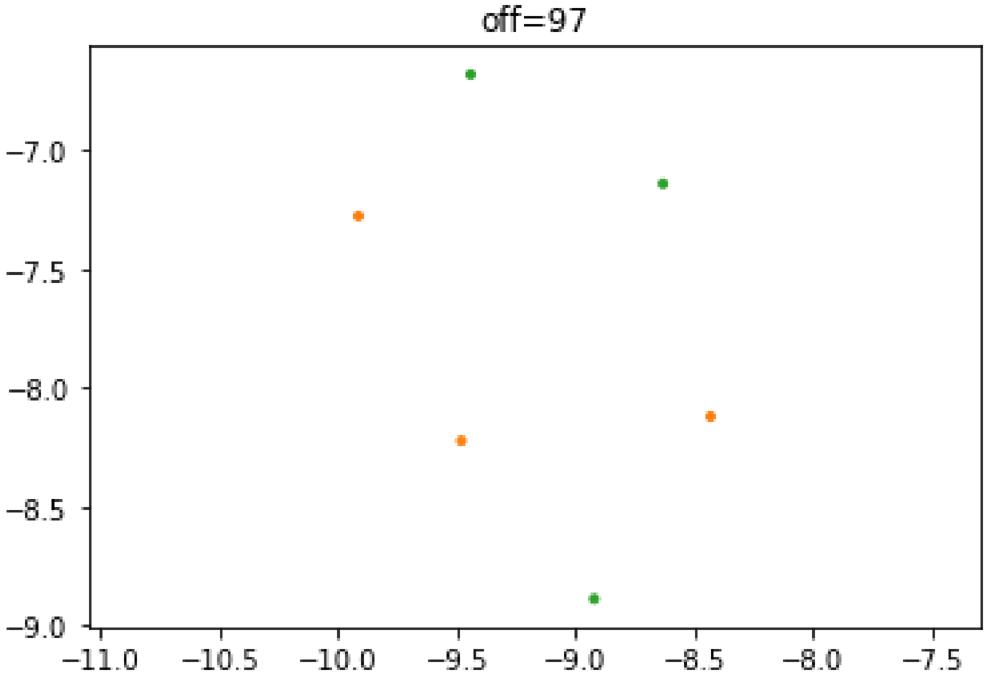
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



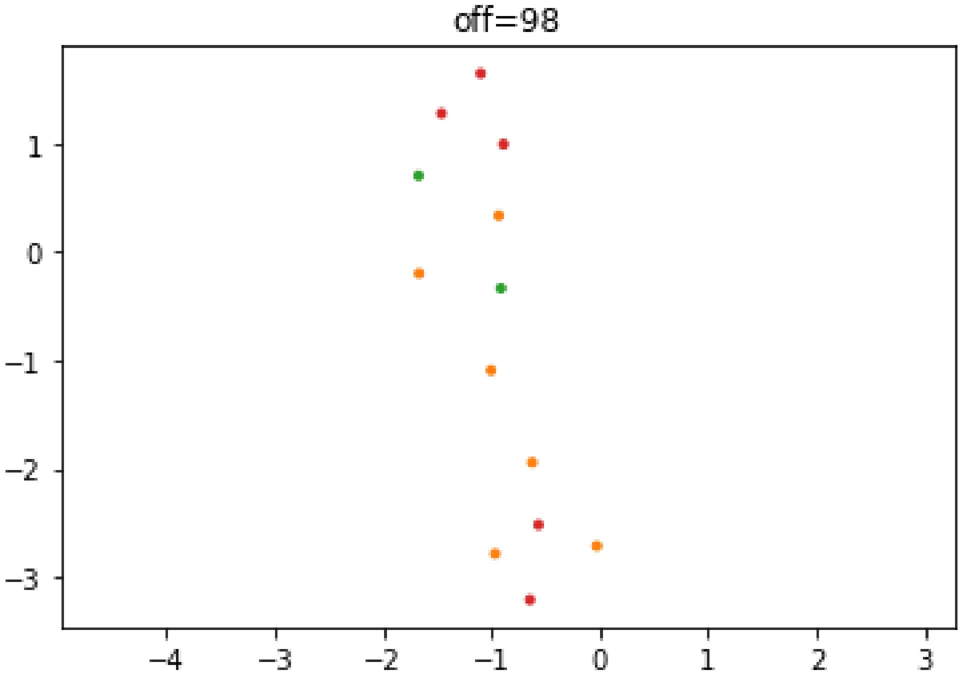
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



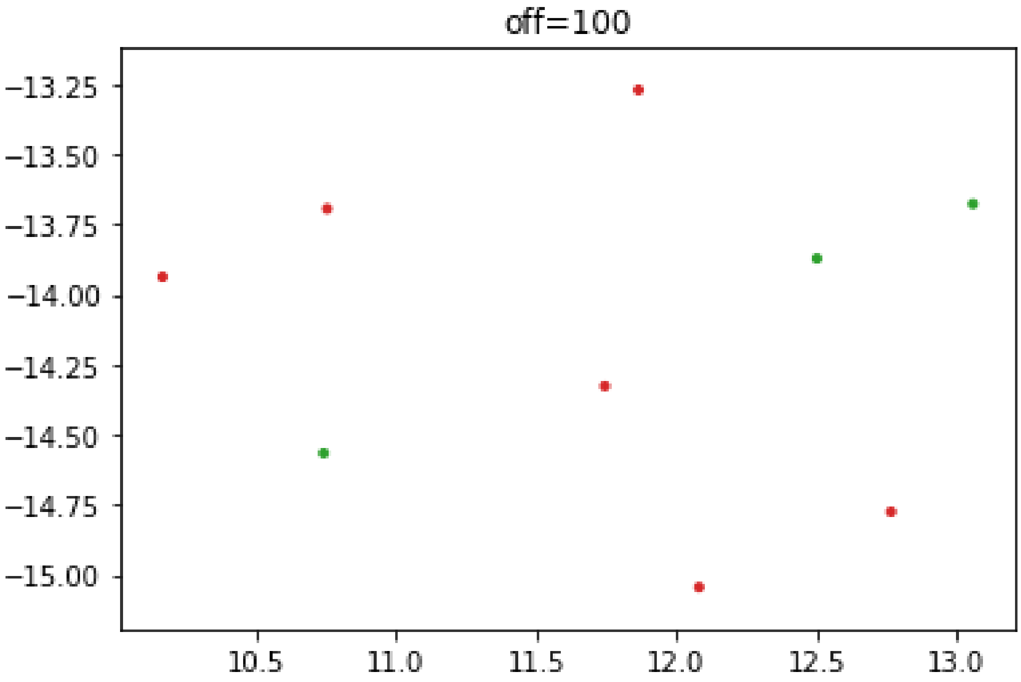
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



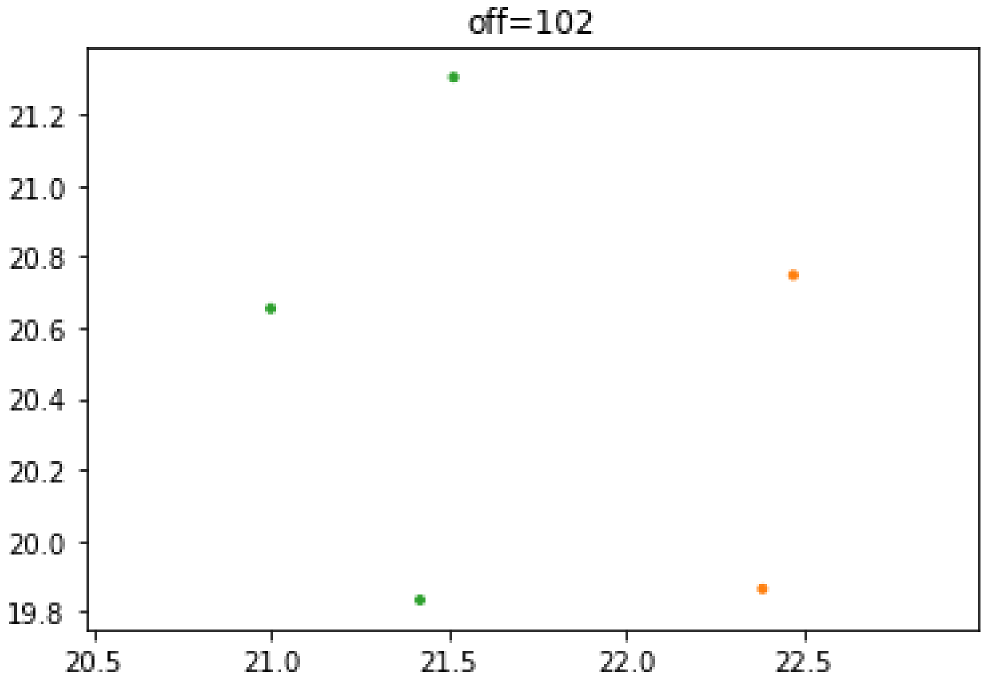
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



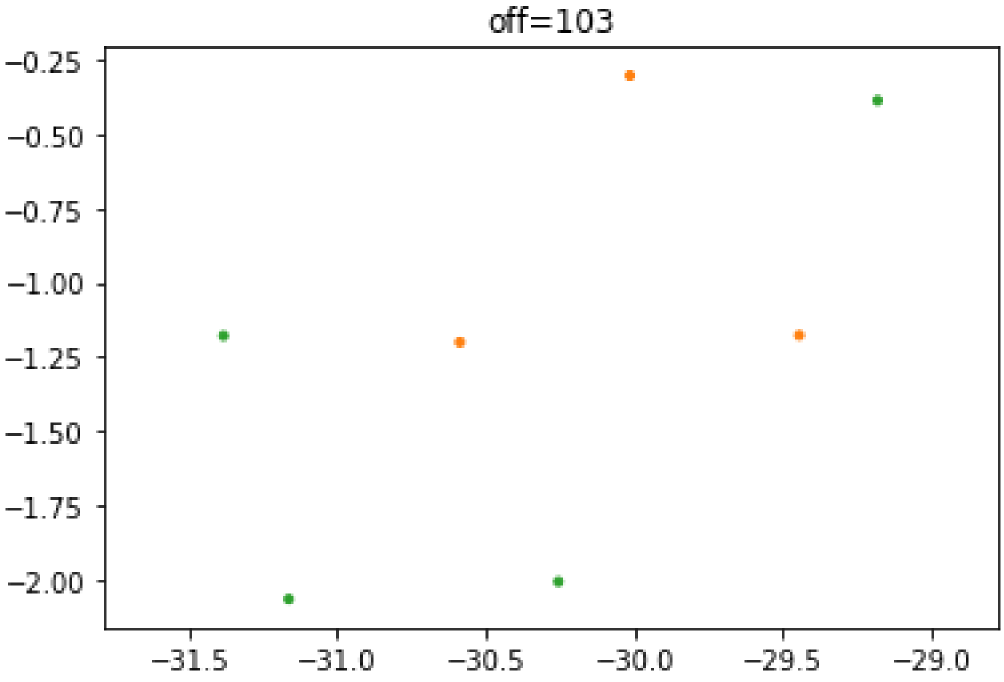
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



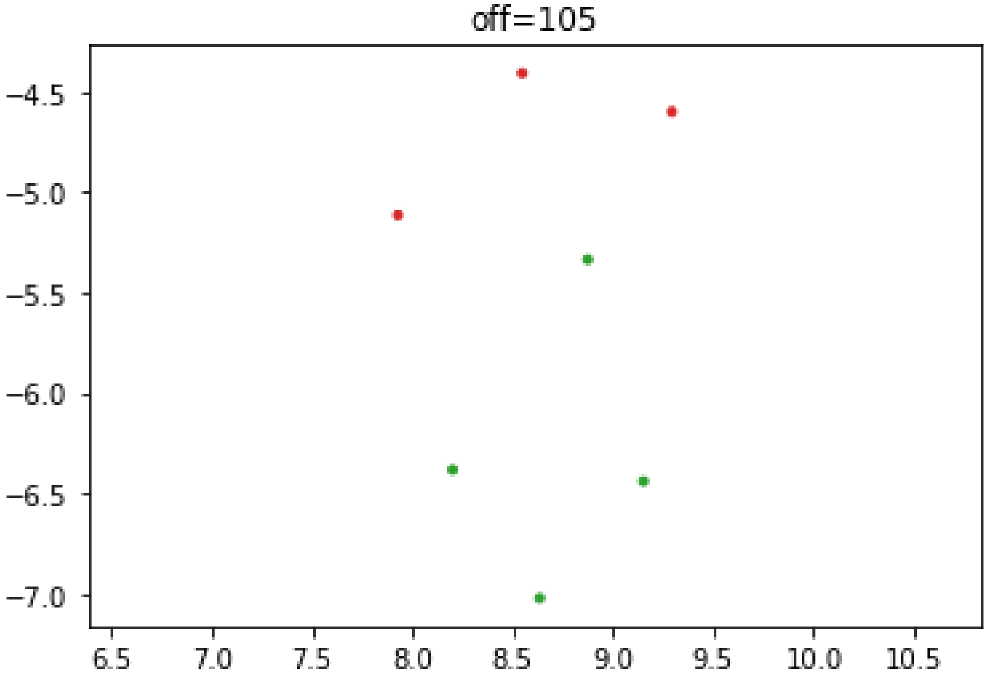
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



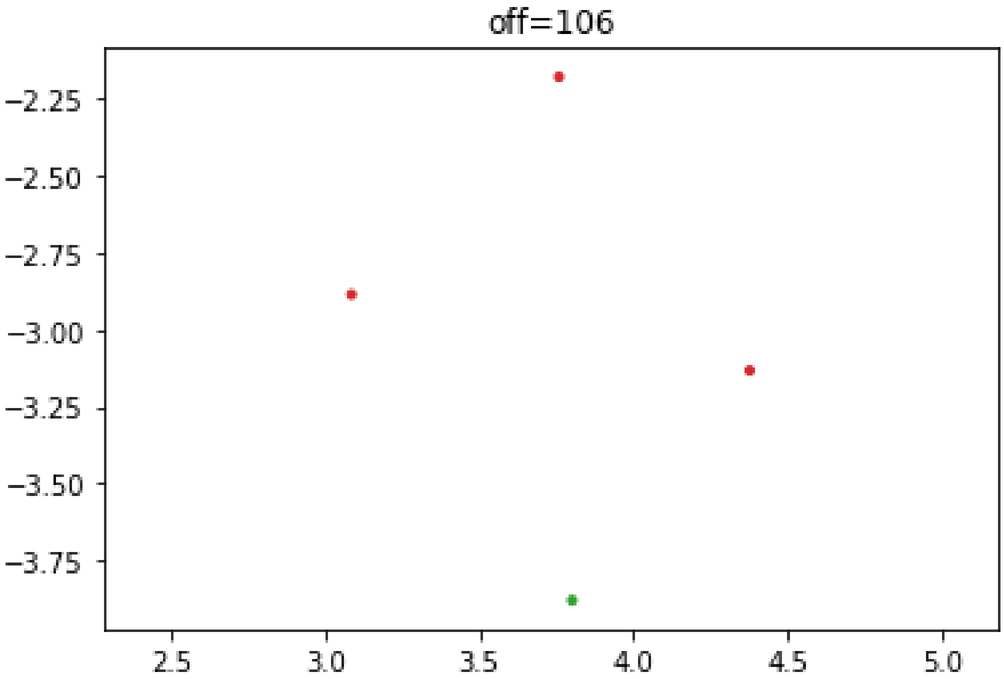
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



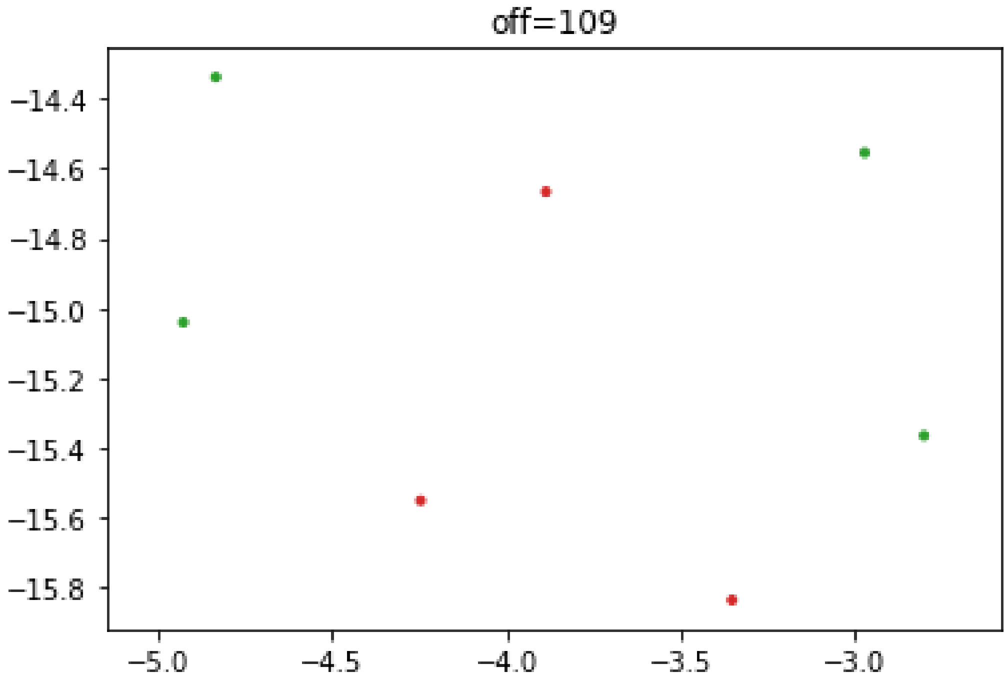
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



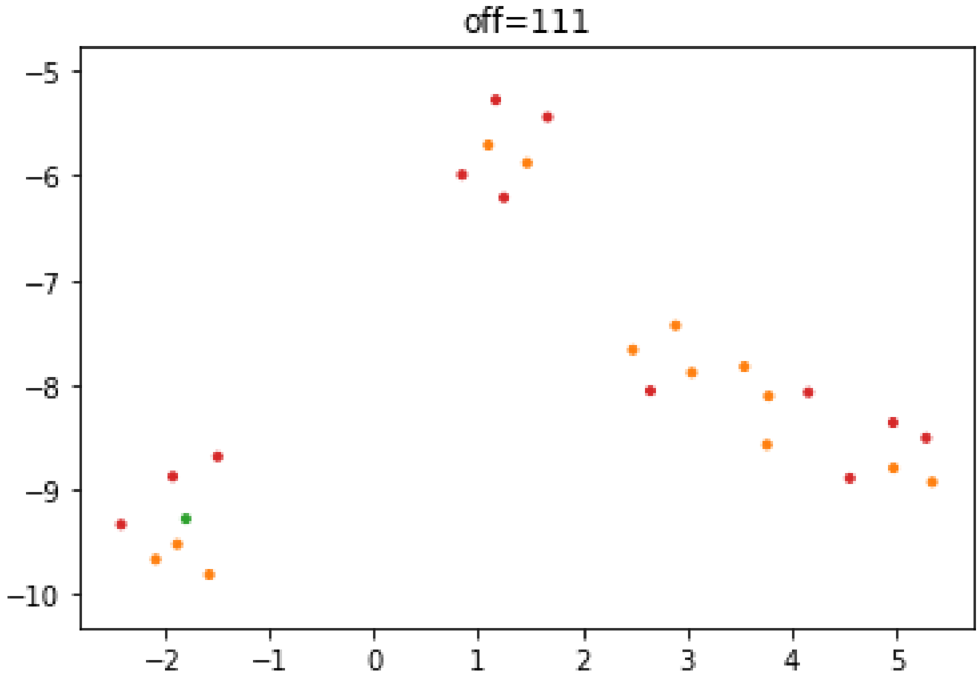
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



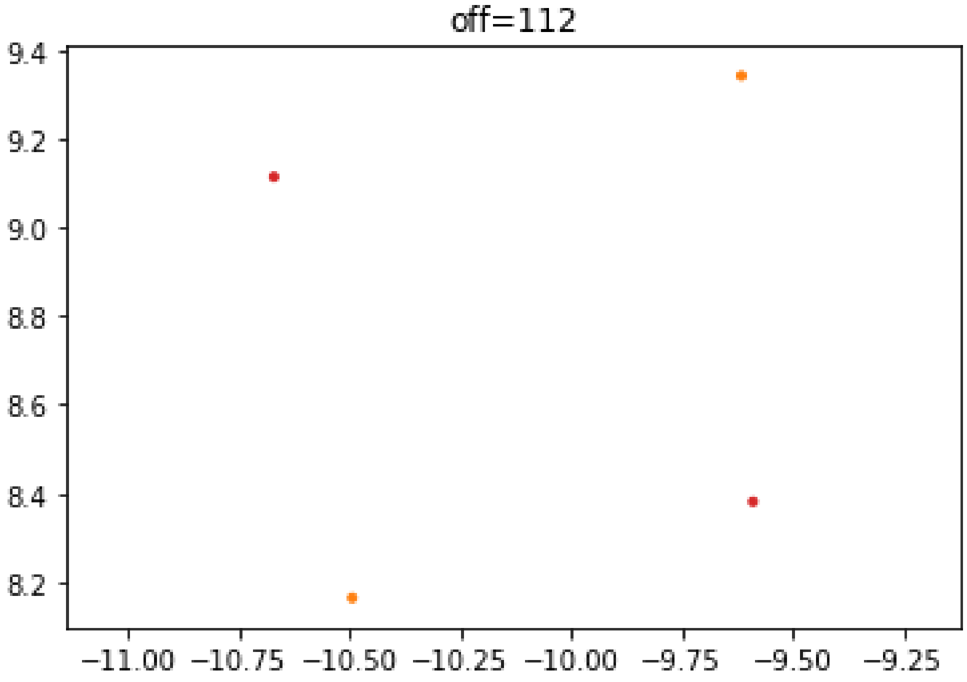
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



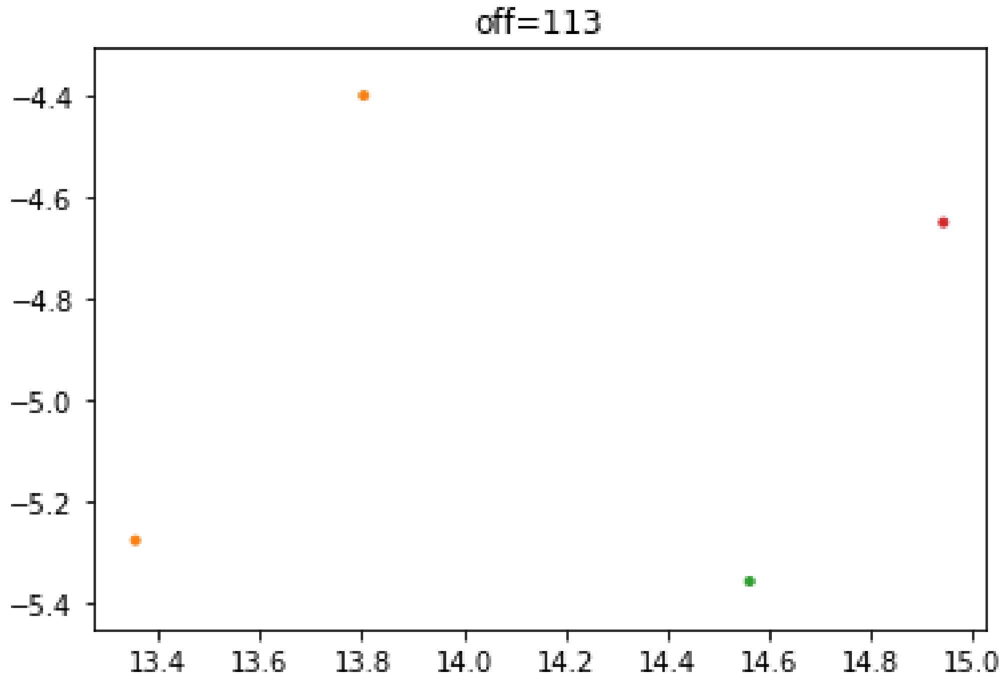
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



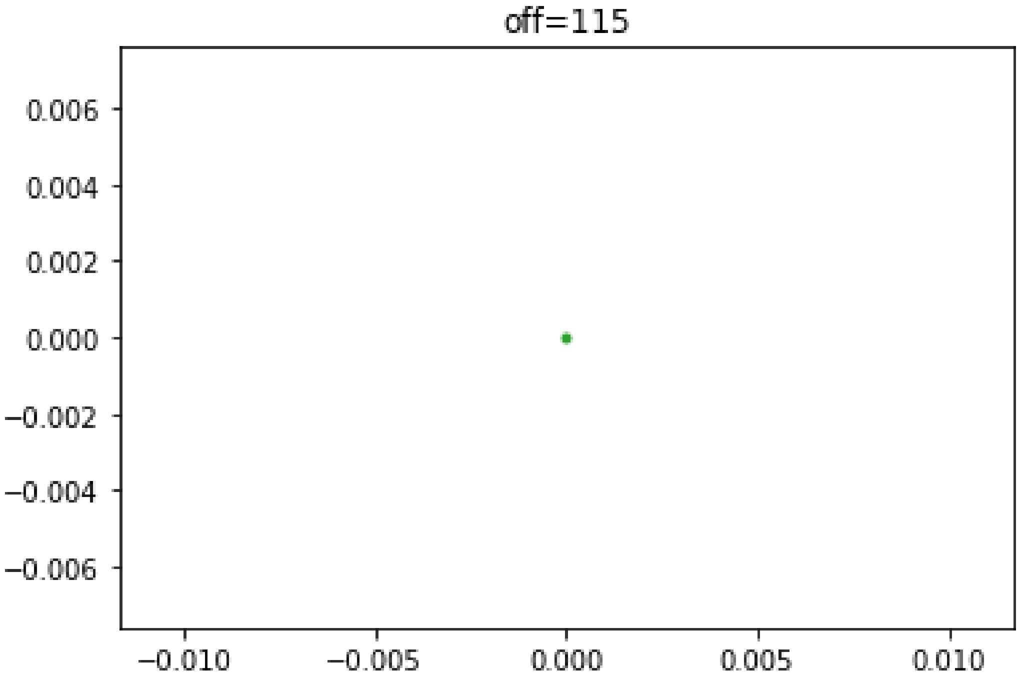
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



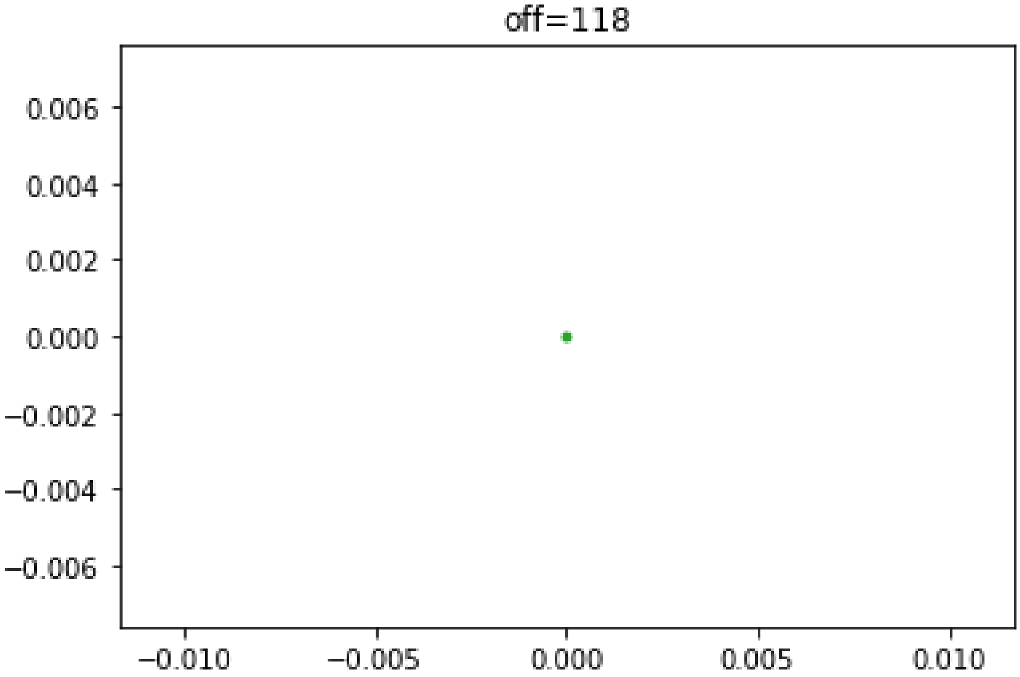
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



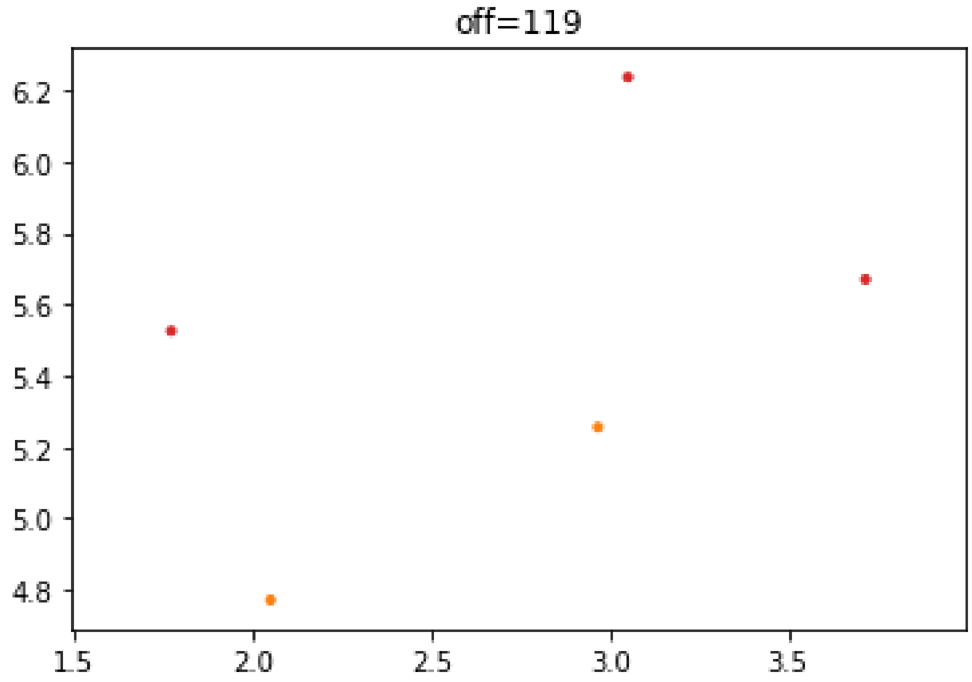
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



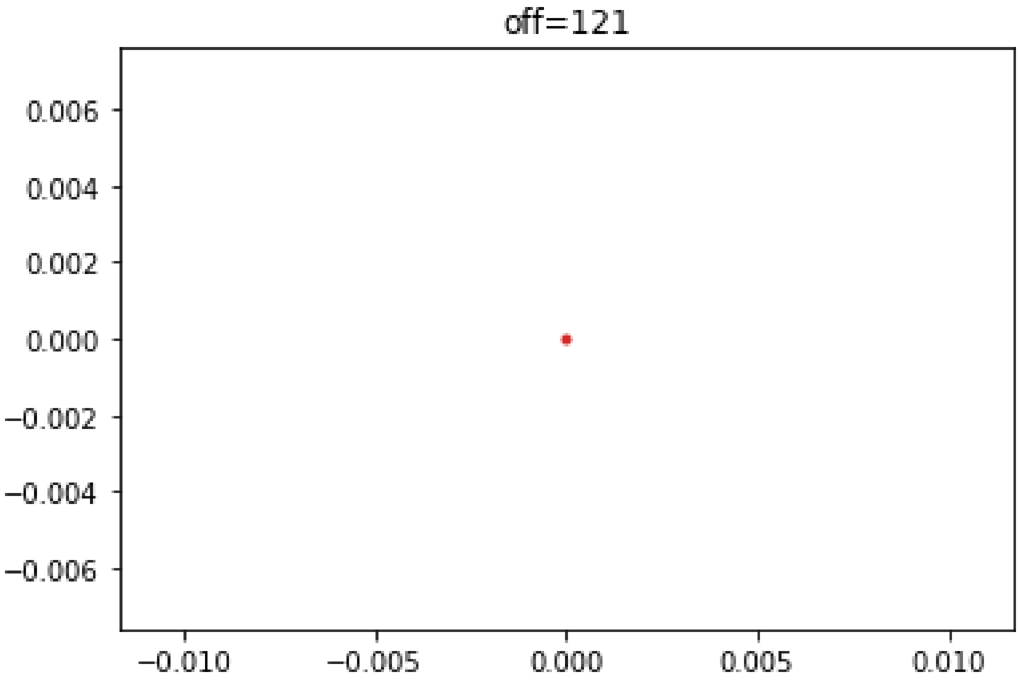
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



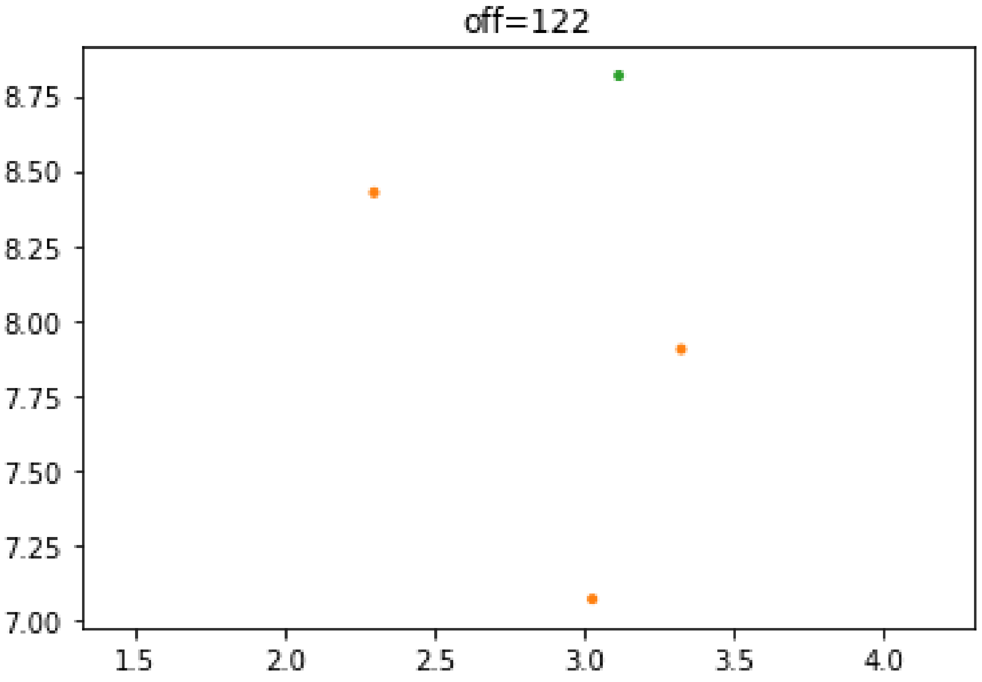
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



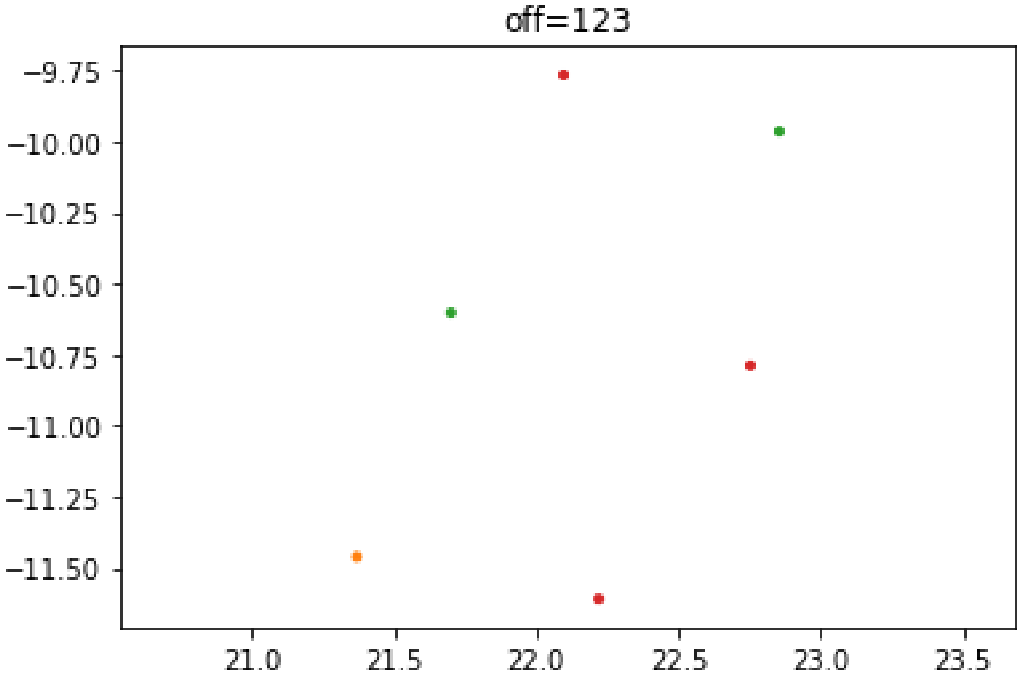
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592

k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
```

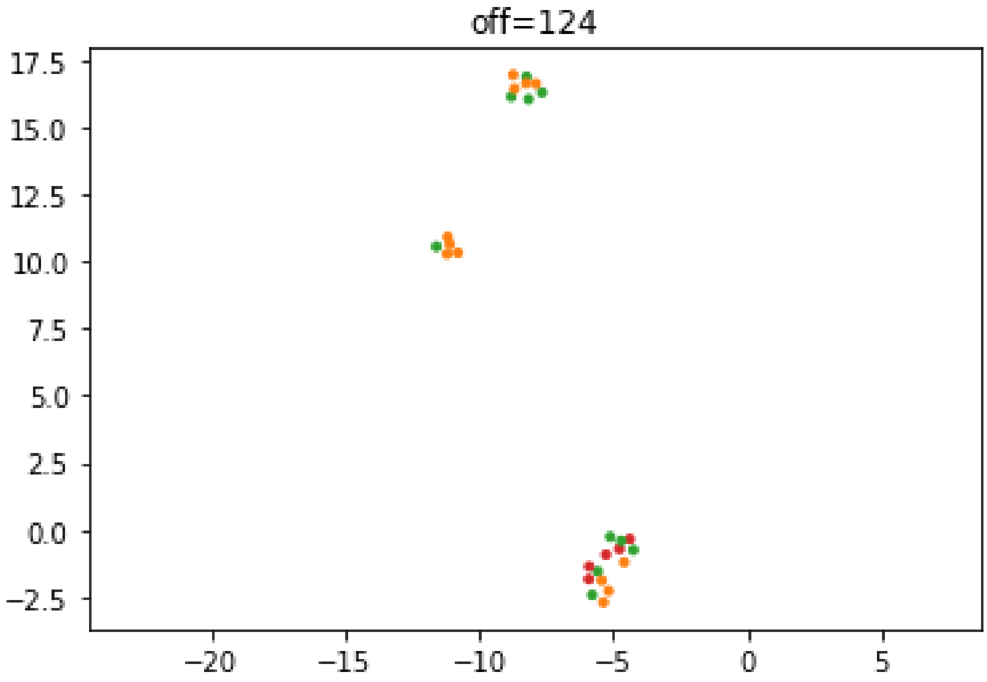


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

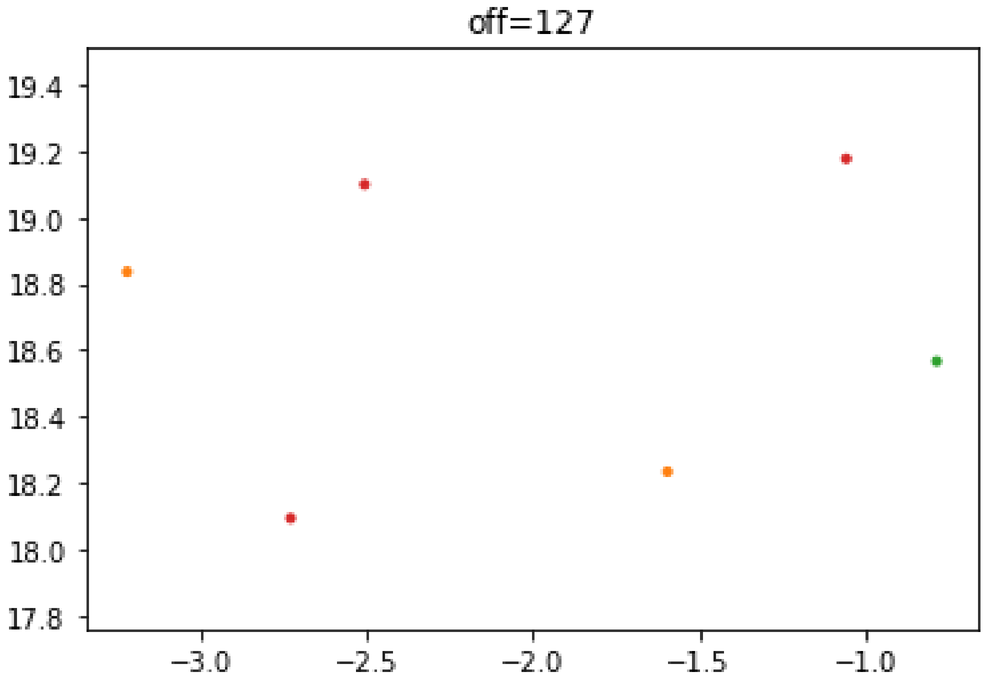


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

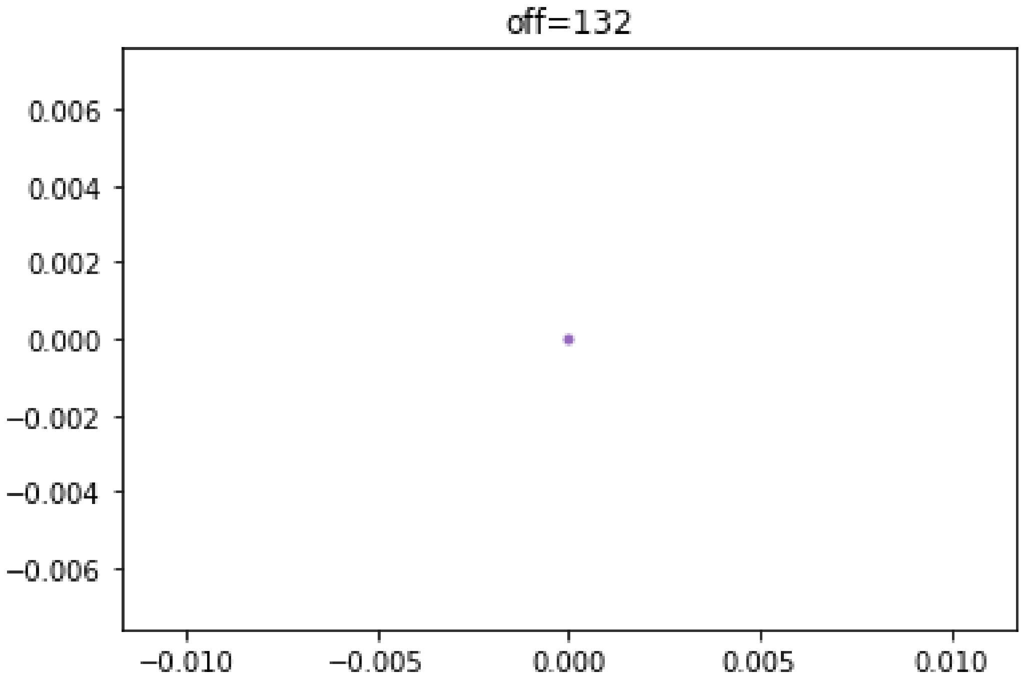




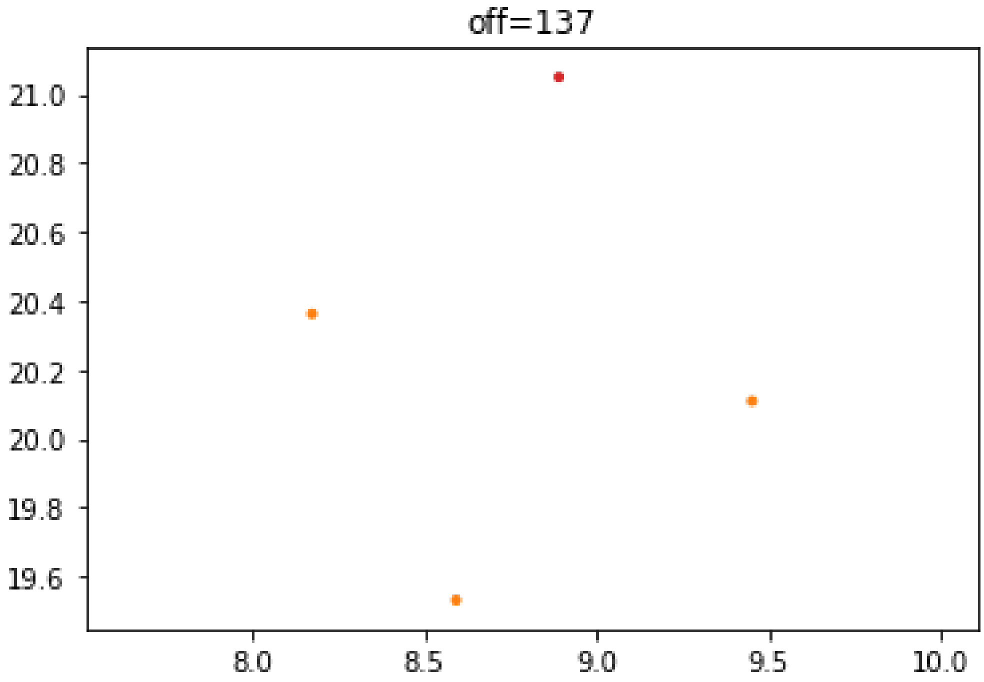
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

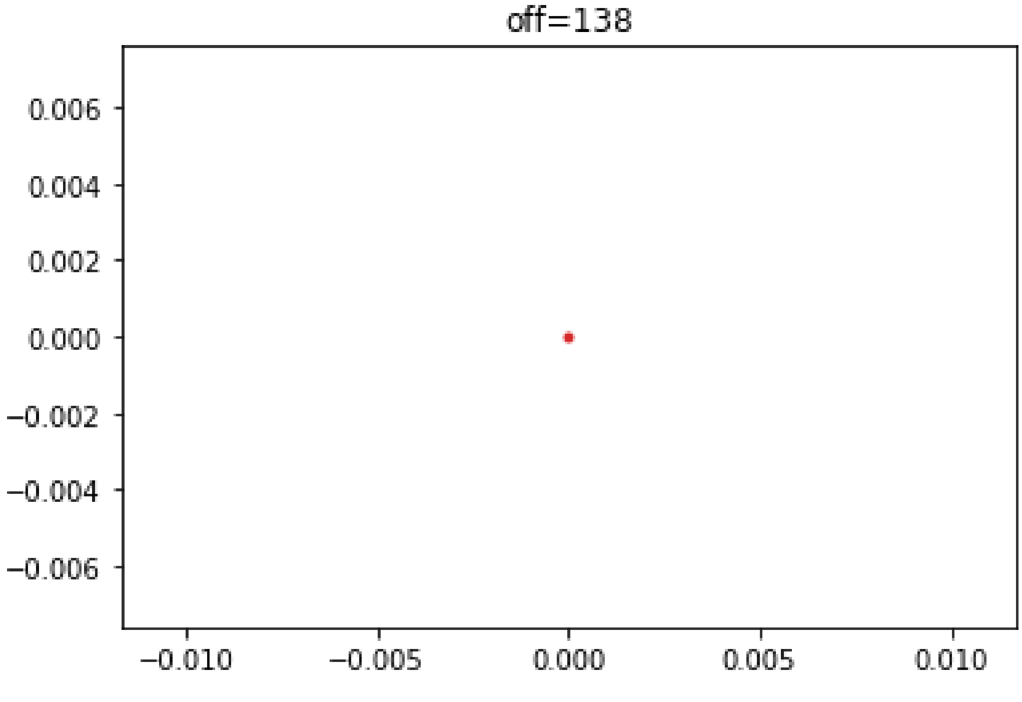


```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
```



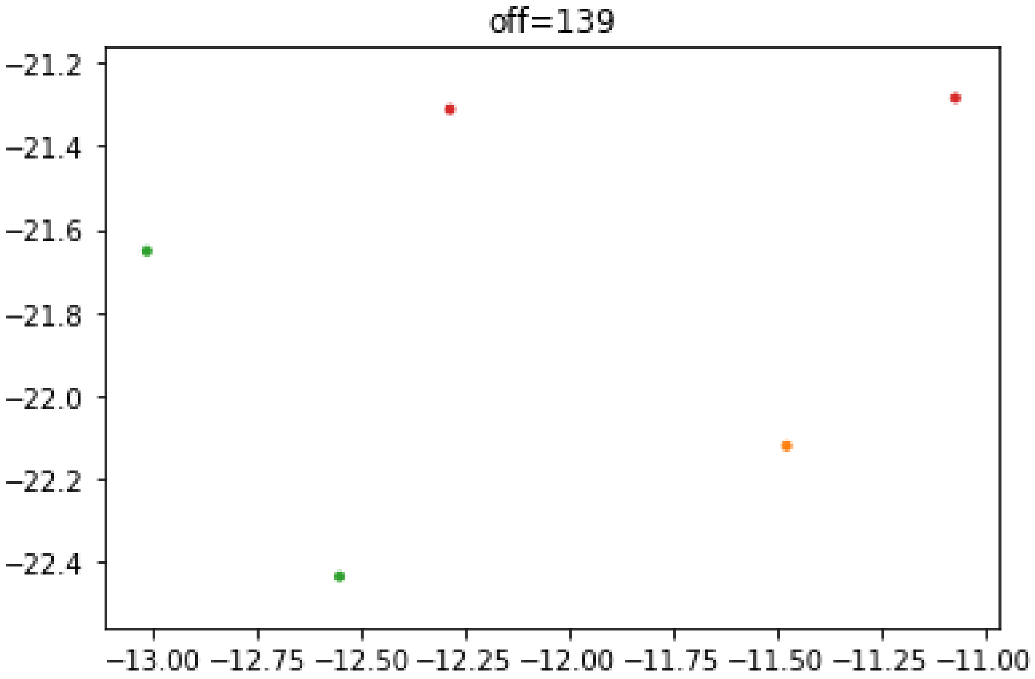
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592

k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.

/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```





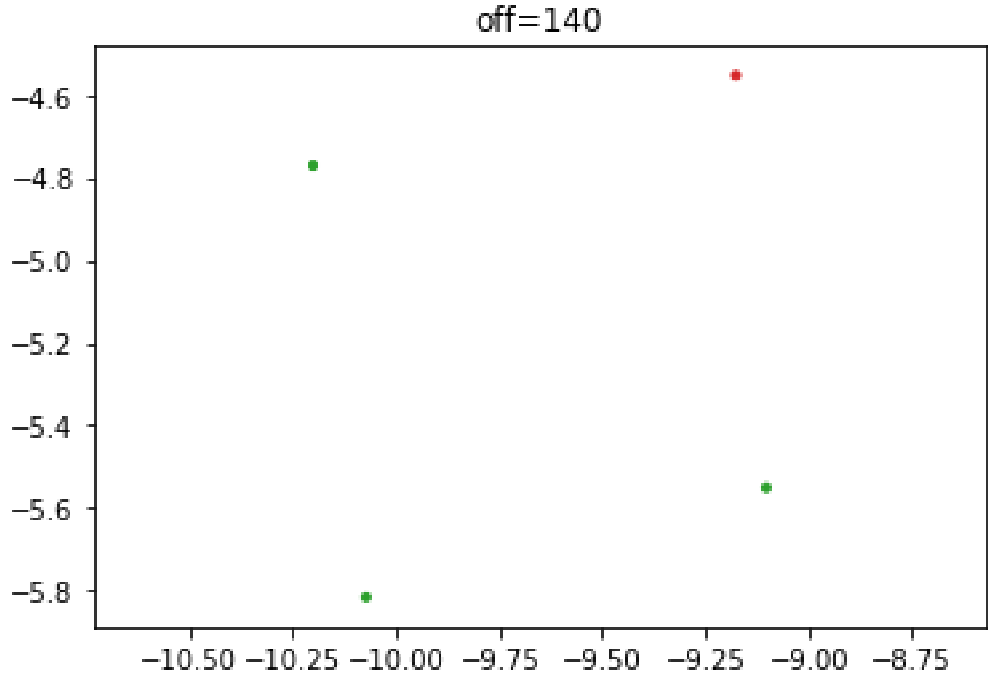
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

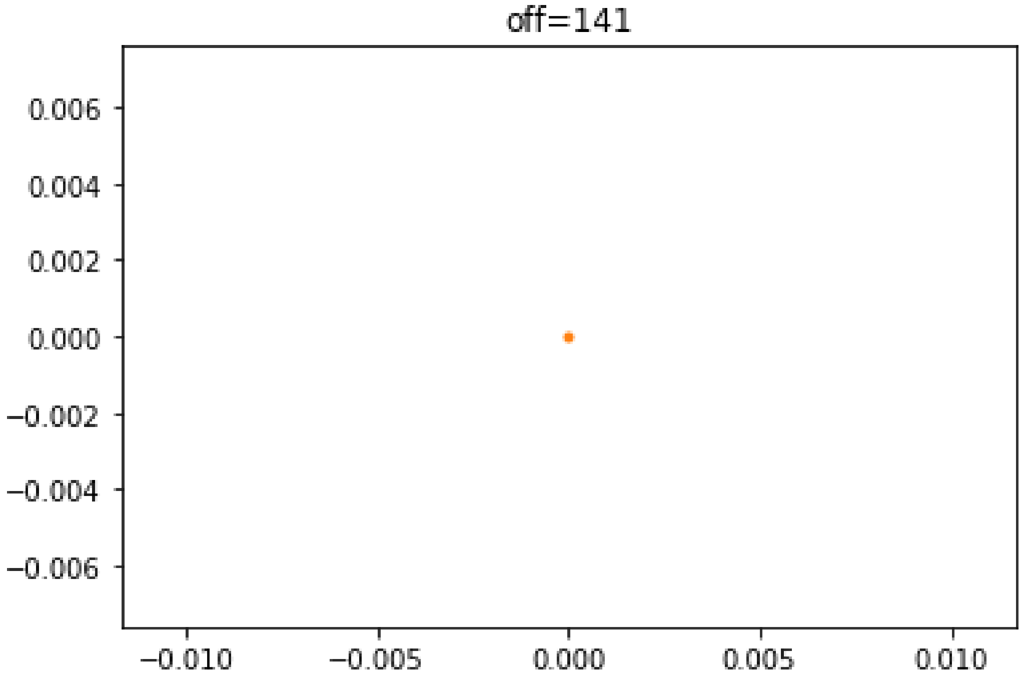
```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



off=139

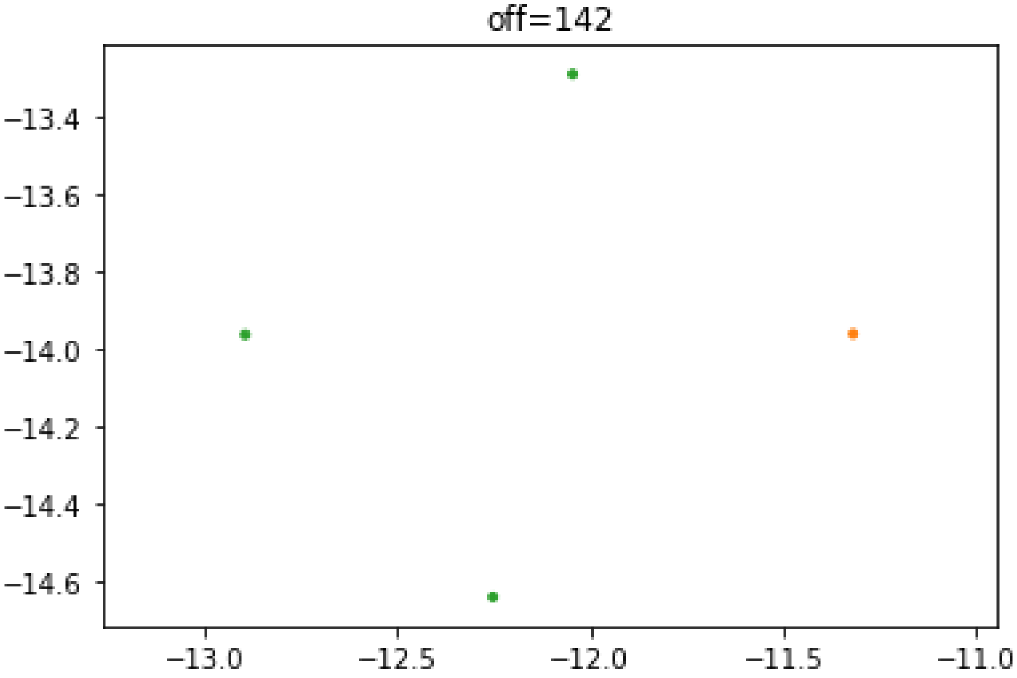
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```



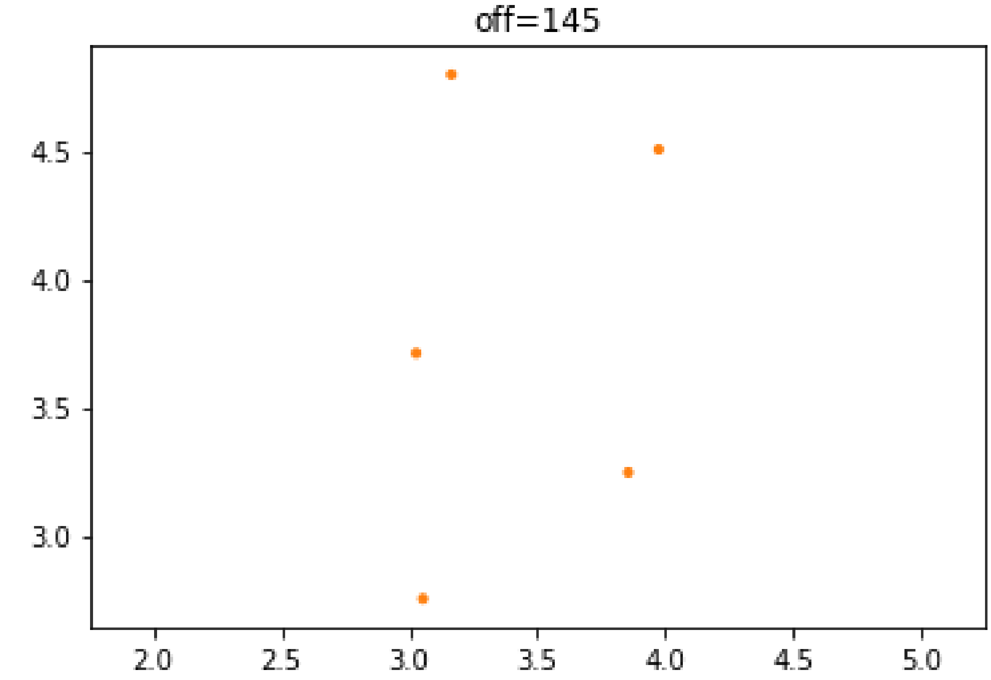


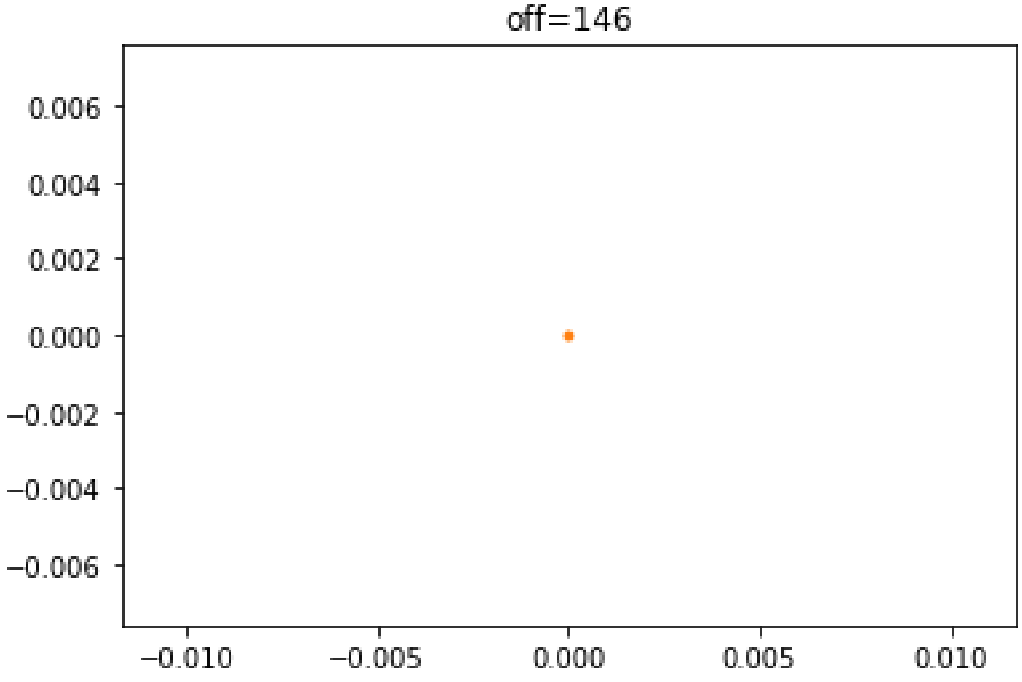
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

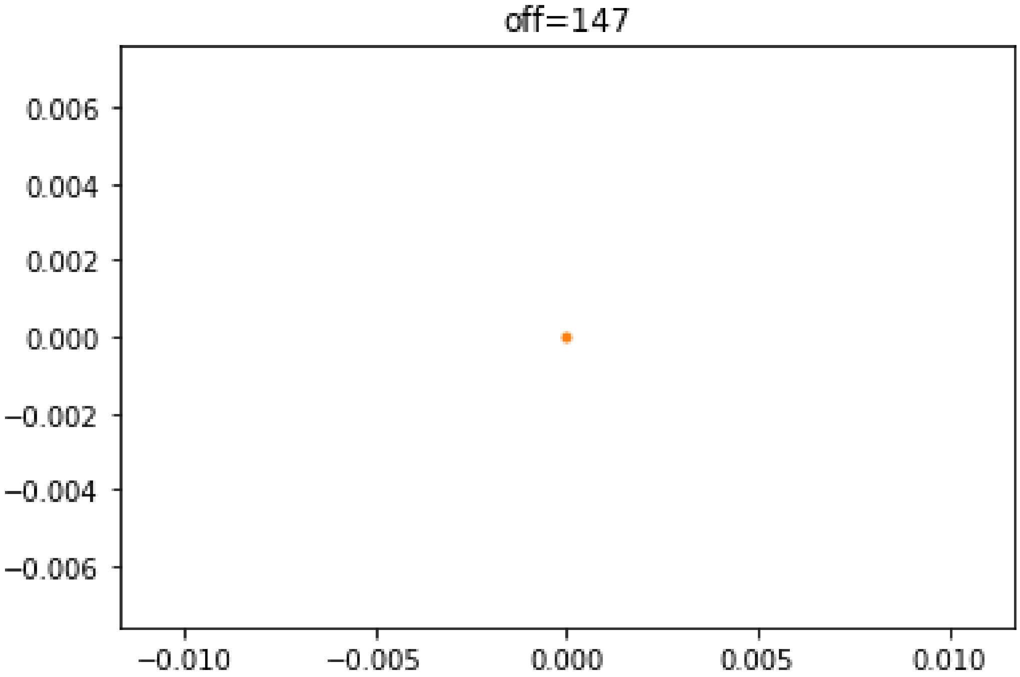
```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

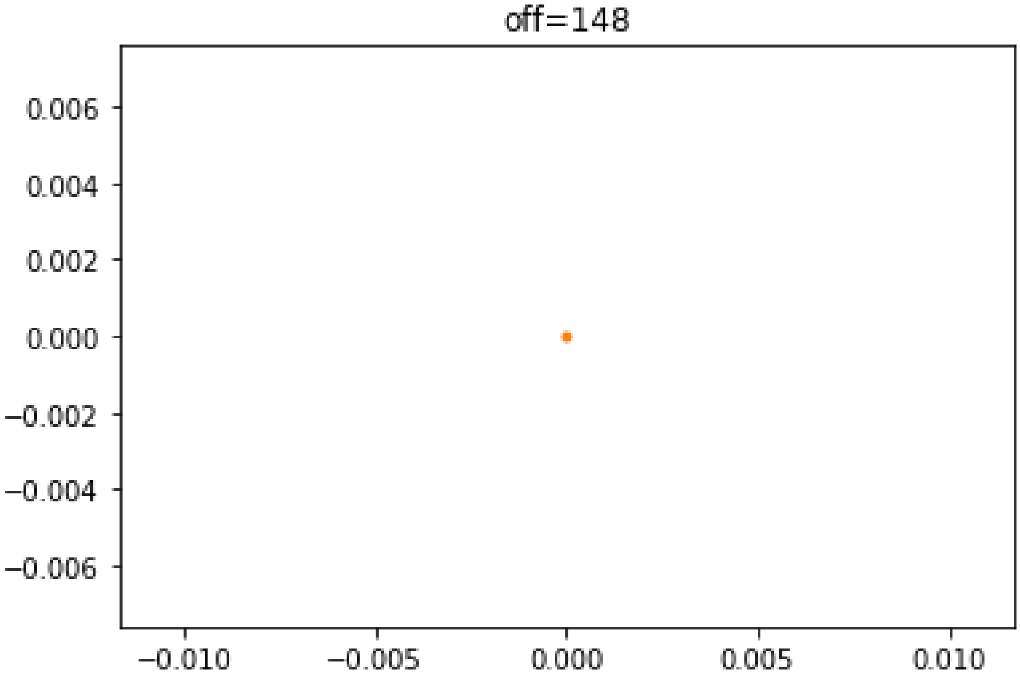


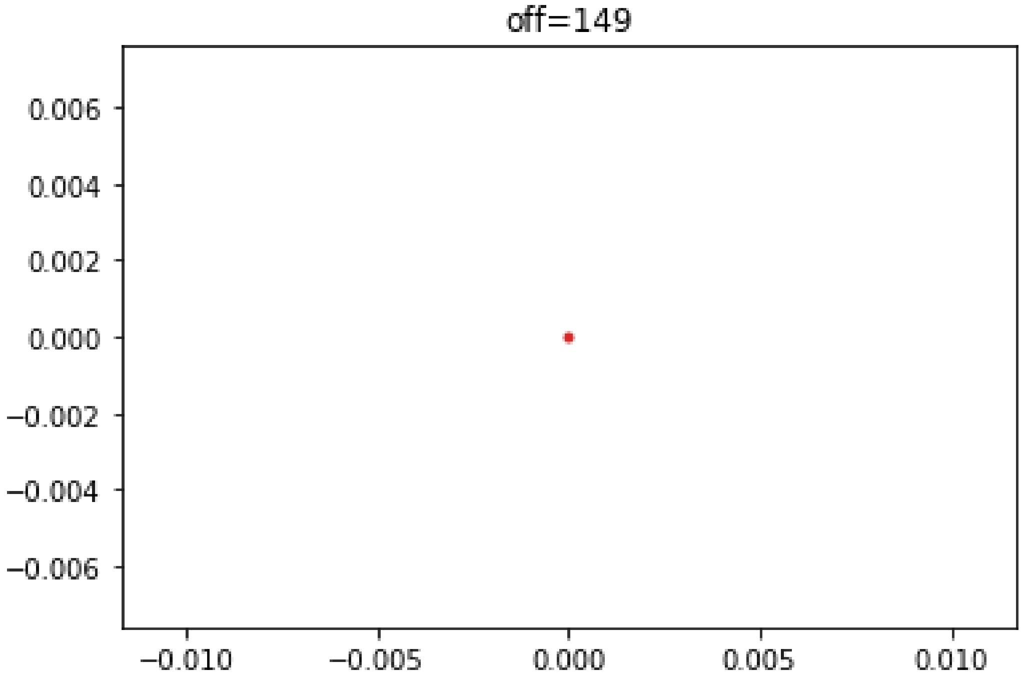
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592
k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

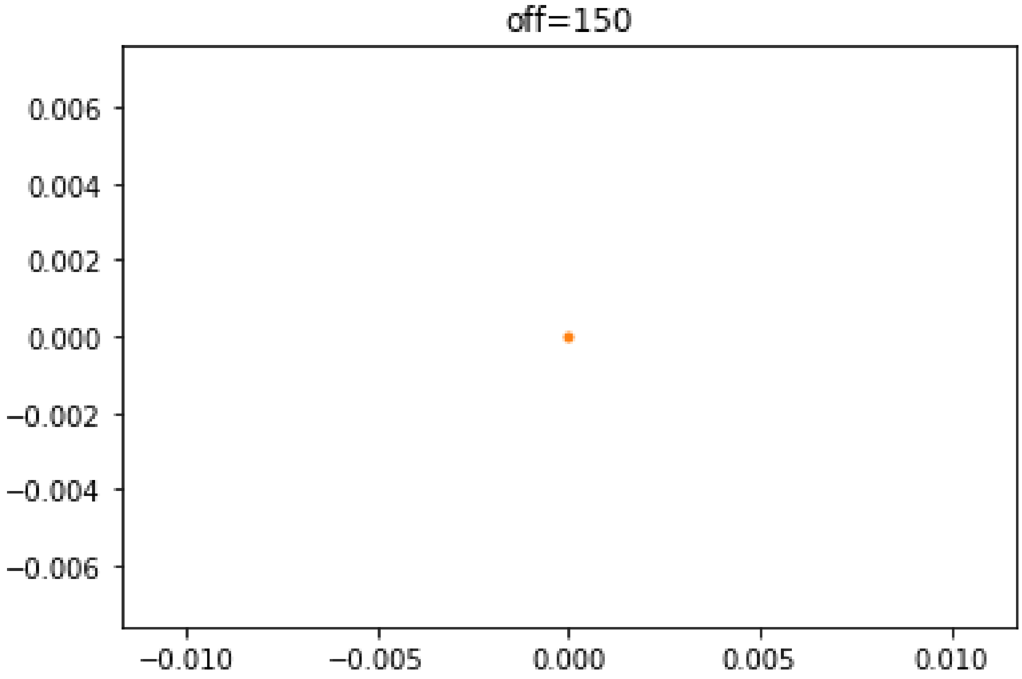

off=145
4.5
4.0
3.5
3.0
2.0
2.5
3.0
3.5
4.0
4.5
5.0


off=146
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010


off=147
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

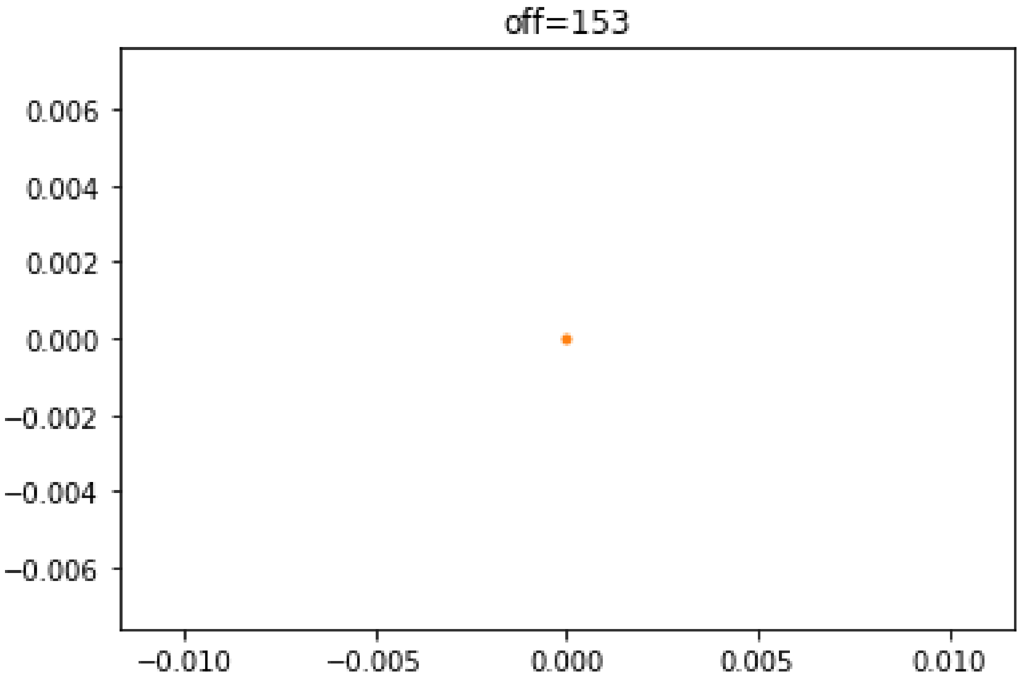

off=148
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

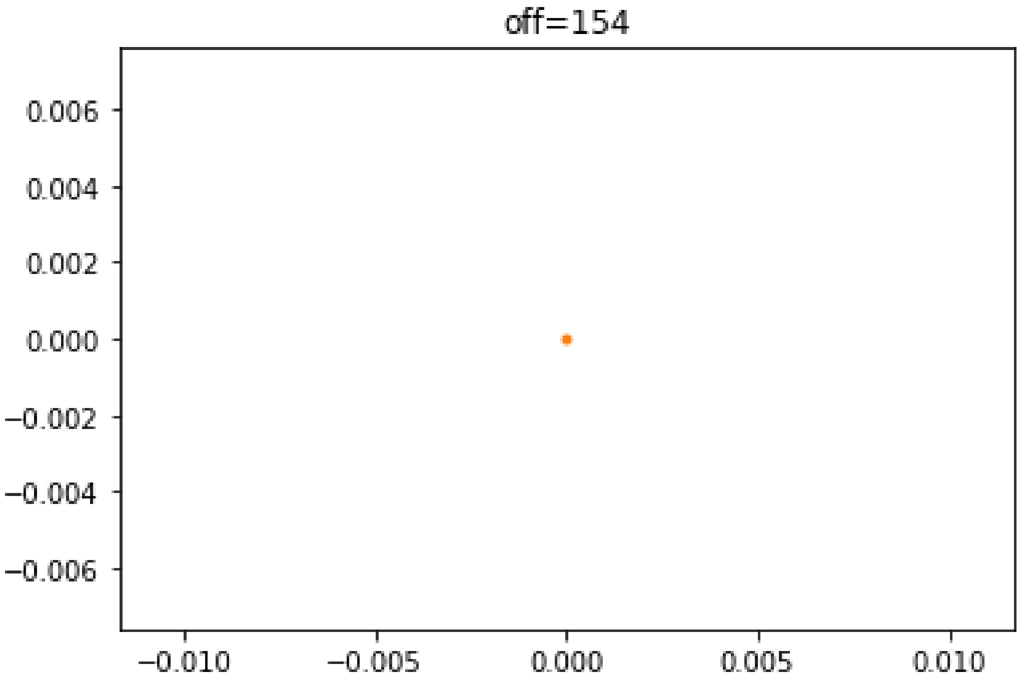




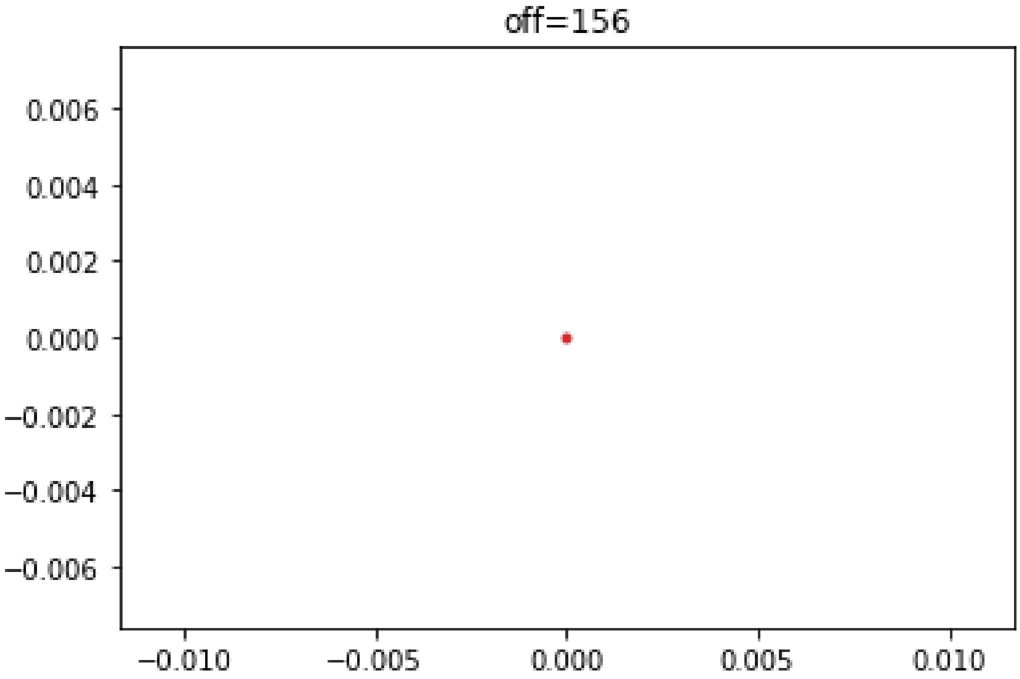
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
```

```
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1

/home/ethan/.local/lib/python3.7/site-packages/scipy/sparse/linalg/eigen/arpack/arpack.py:1592

k >= N for N * N square matrix. Attempting to use scipy.linalg.eigh instead.

/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:

n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

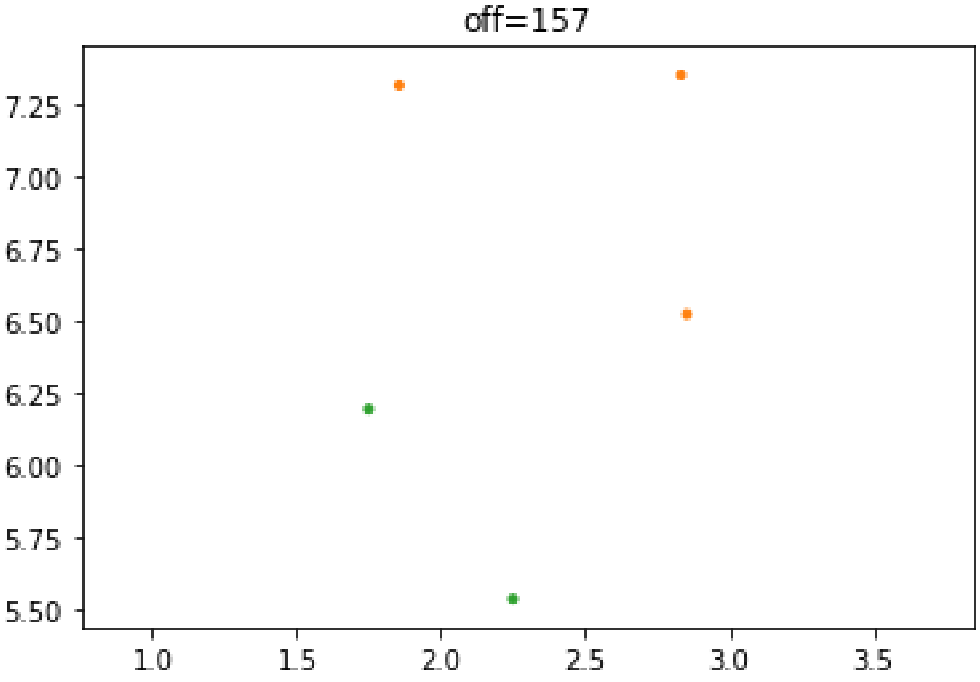




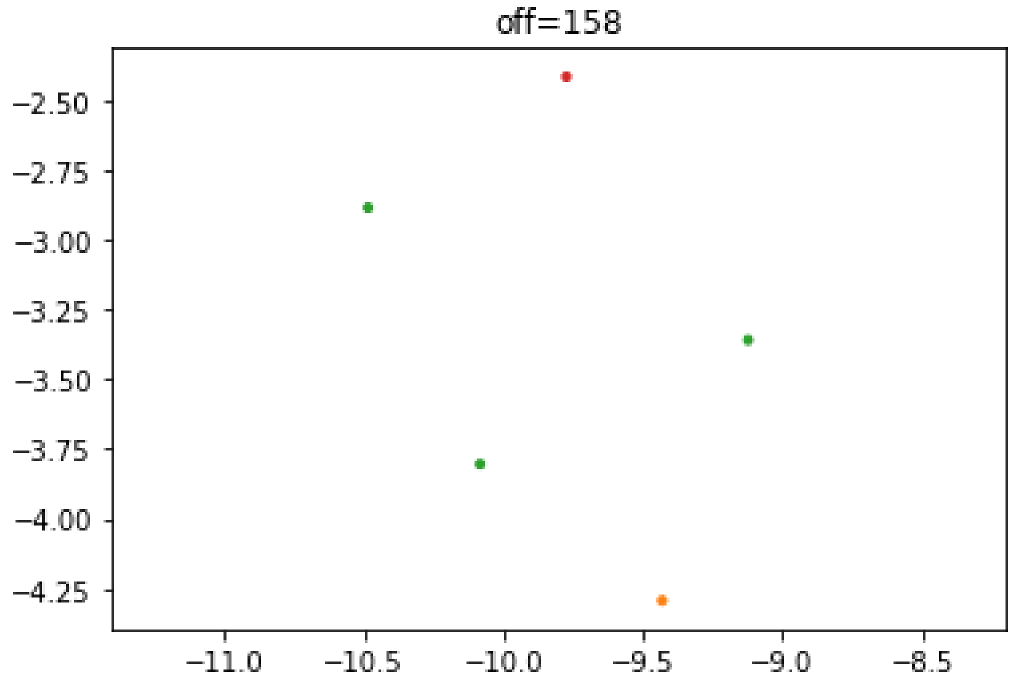
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

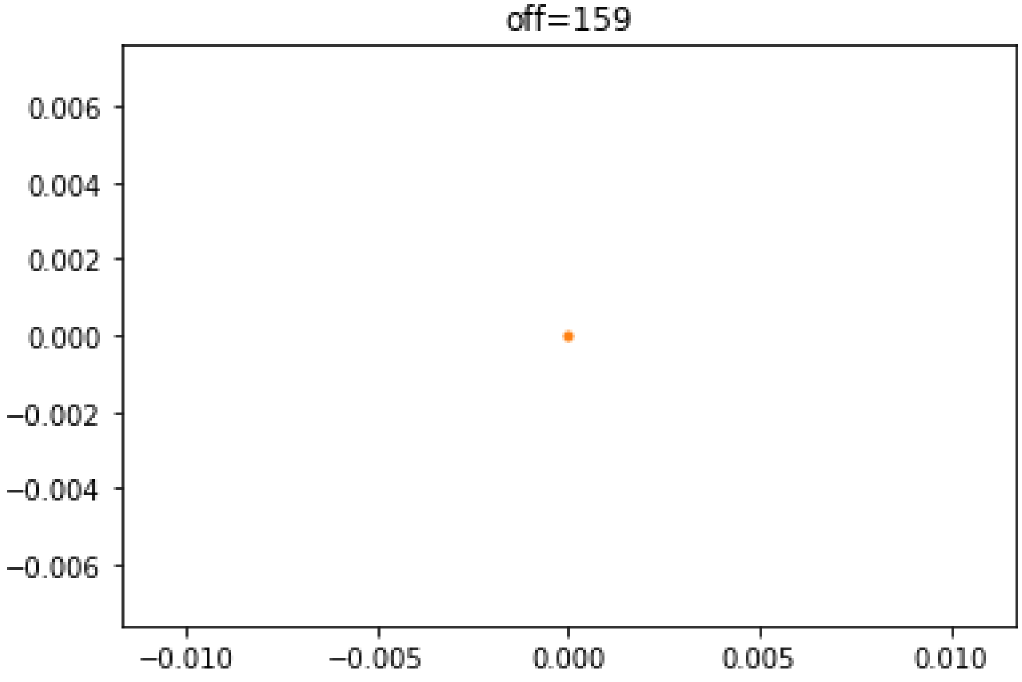


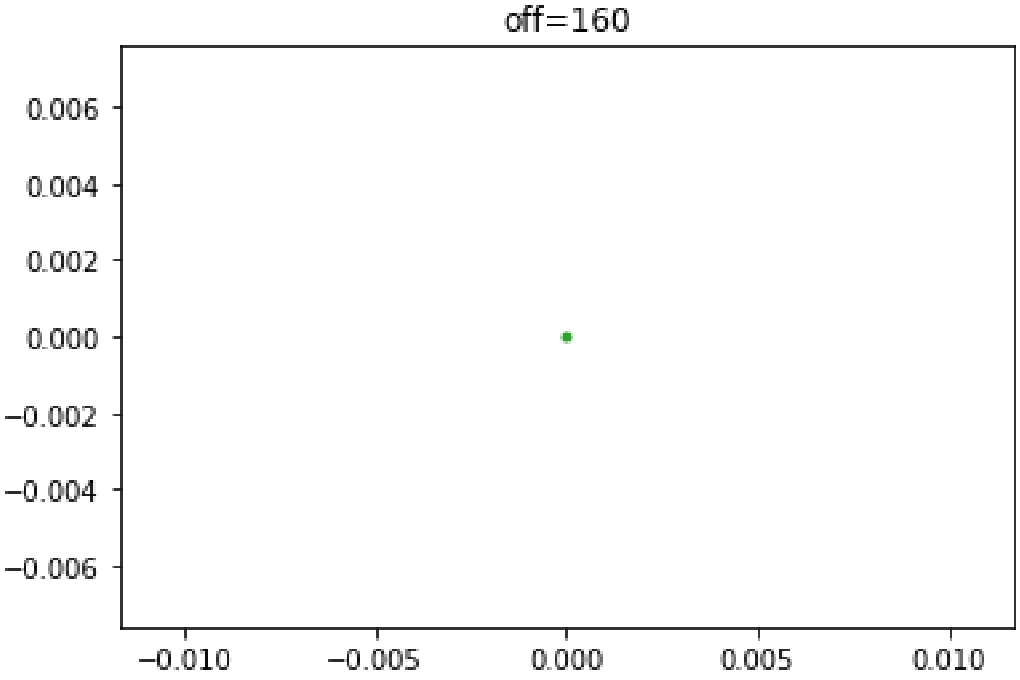
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

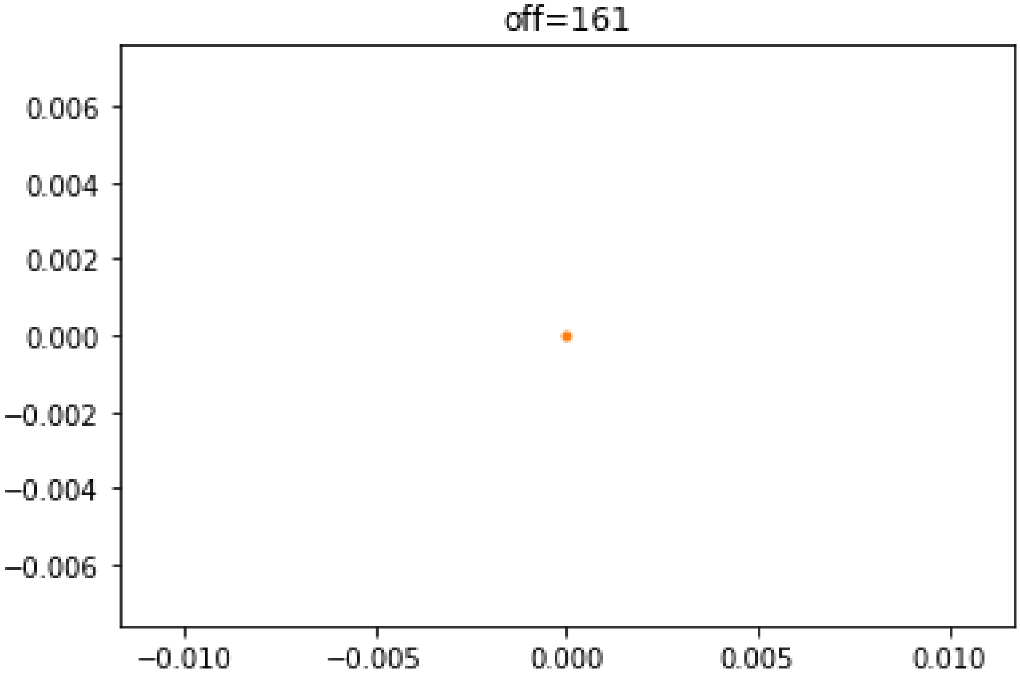


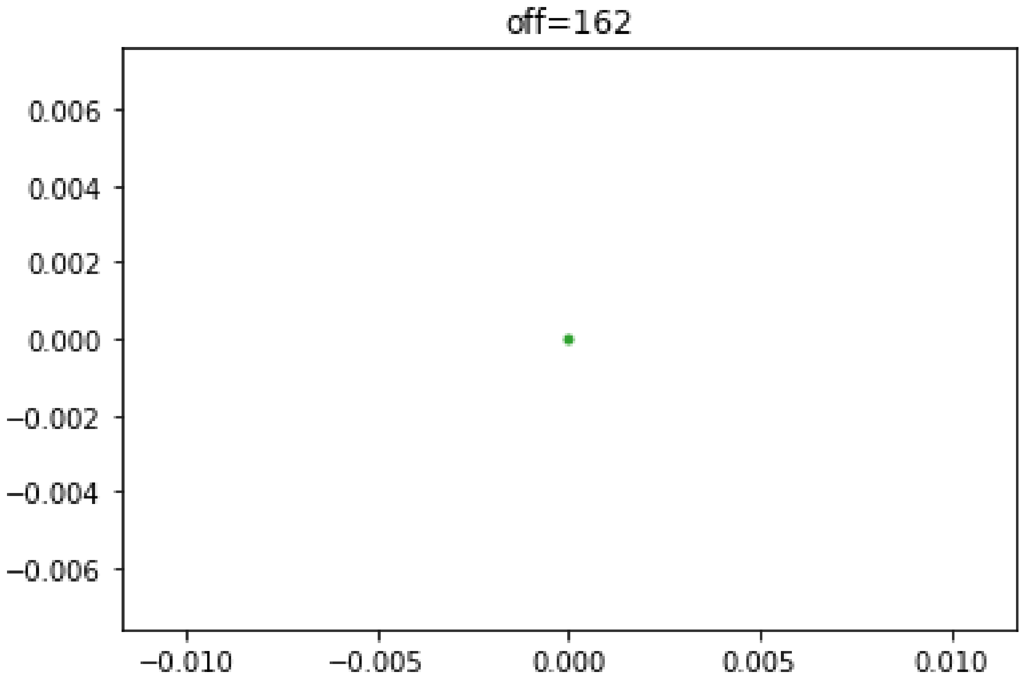
```
/home/ethan/.local/lib/python3.7/site-packages/umap/umap_.py:1383: UserWarning:
n_neighbors is larger than the dataset size; truncating to X.shape[0] - 1
```

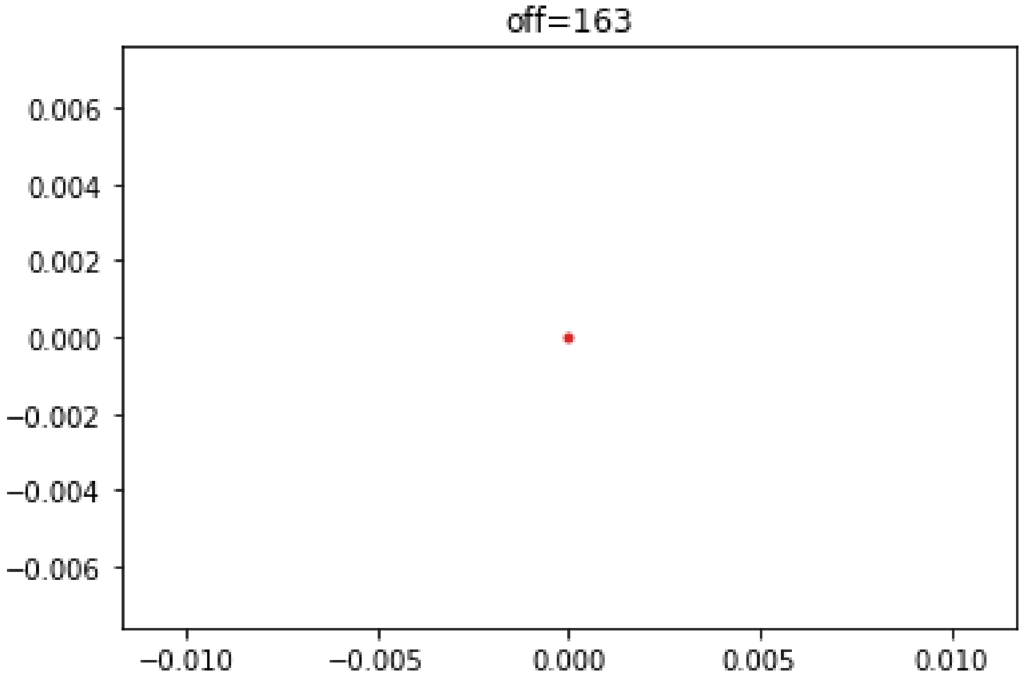

off=158
−2.50
−2.75
−3.00
−3.25
−3.50
−3.75
−4.00
−4.25
−11.0
−10.5
−10.0
−9.5
−9.0
−8.5

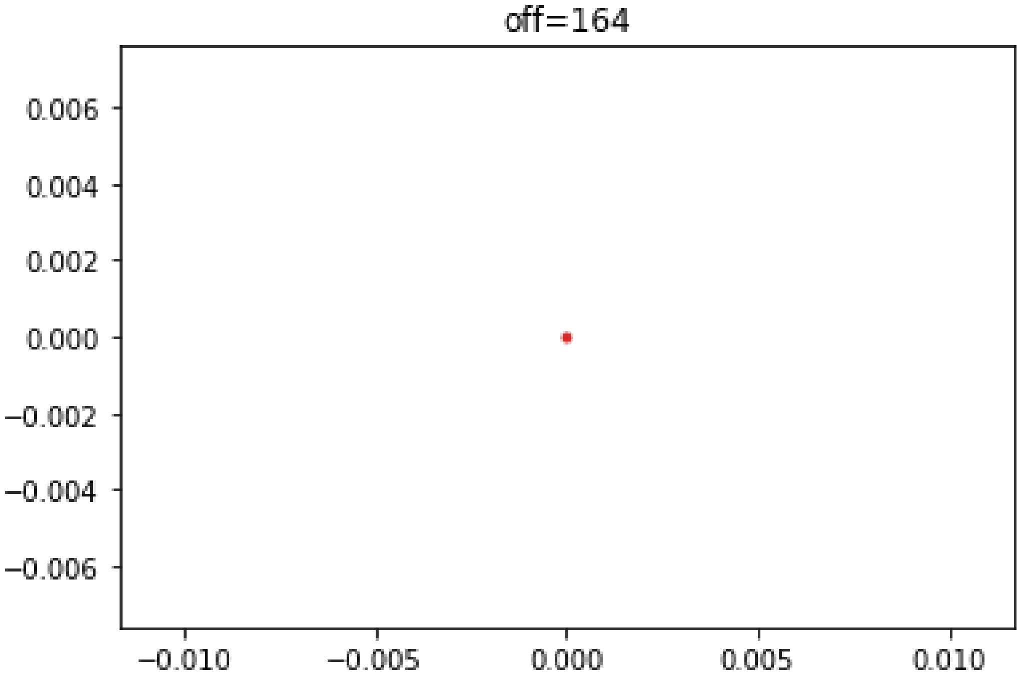

off=159
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

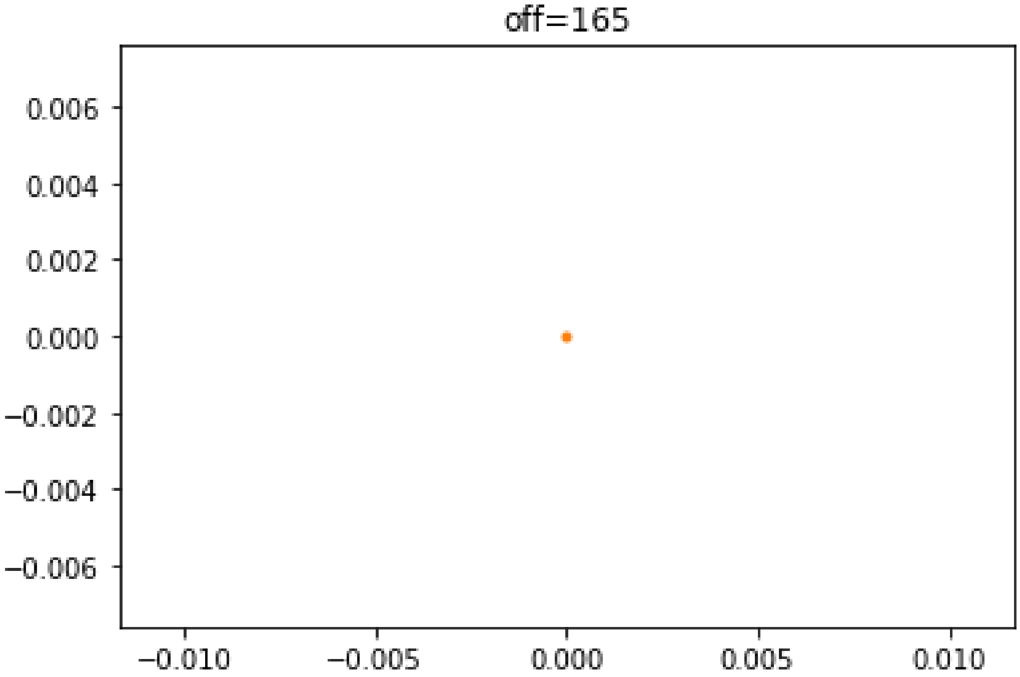

off=160
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

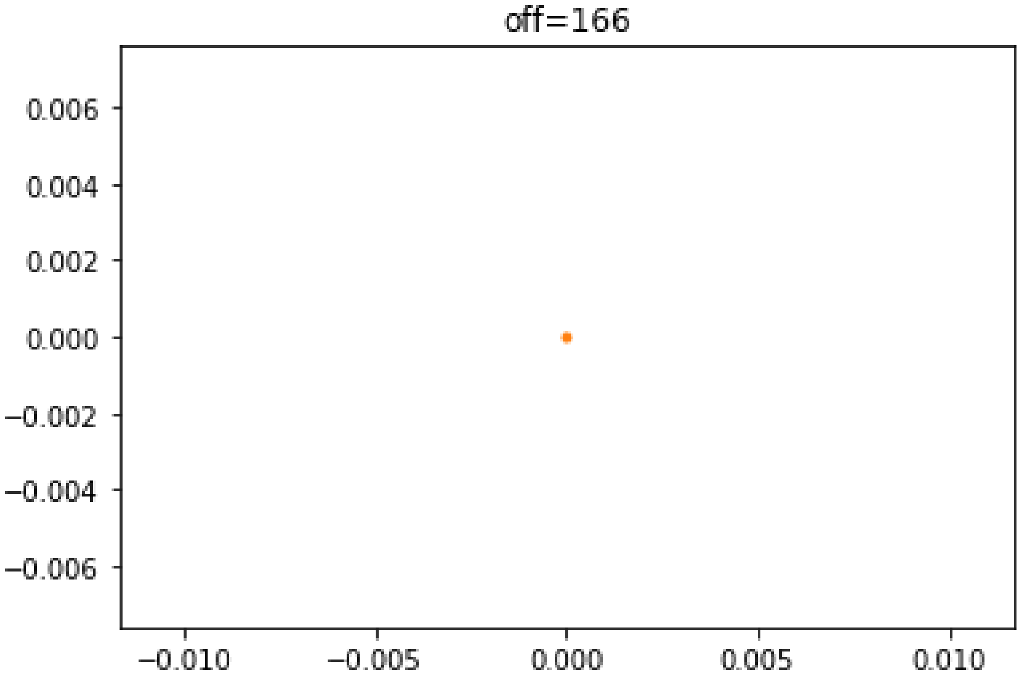

off=161
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

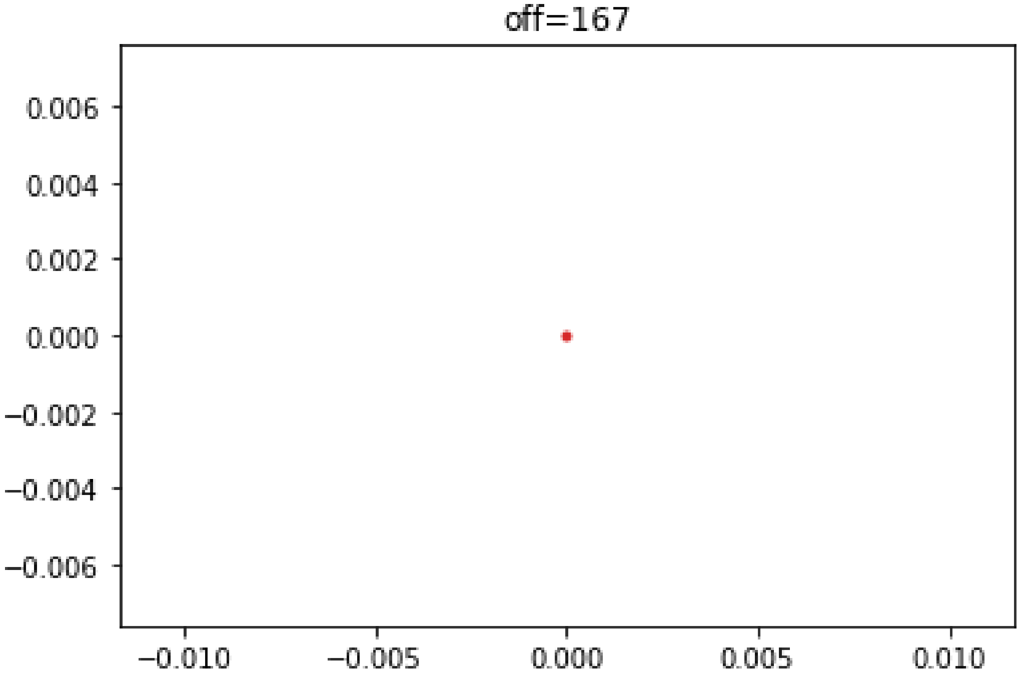

off=162

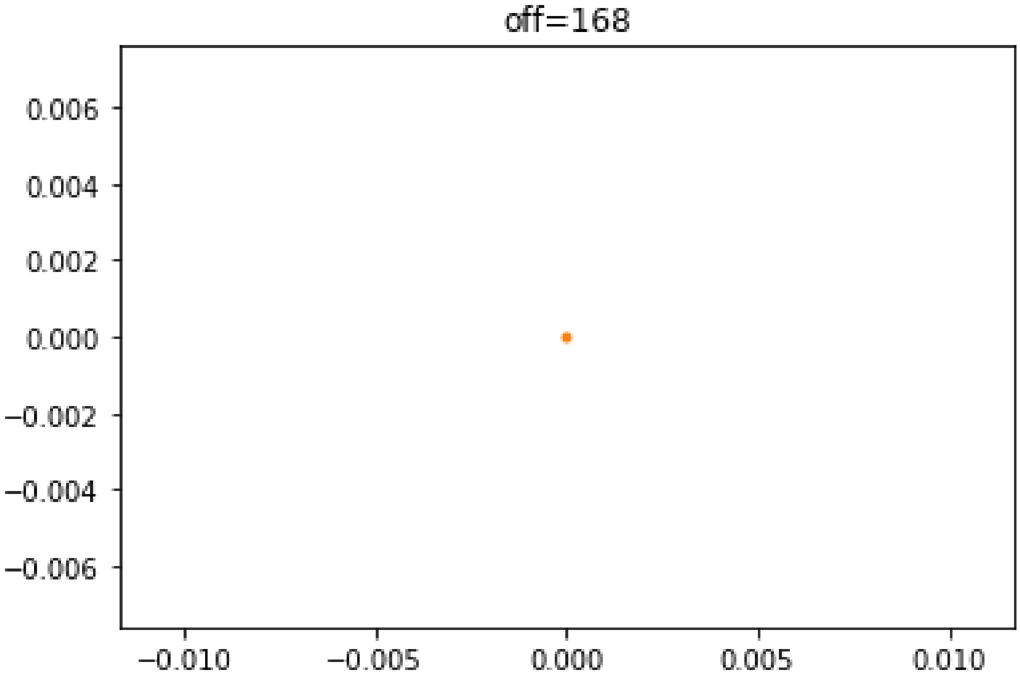

off=163

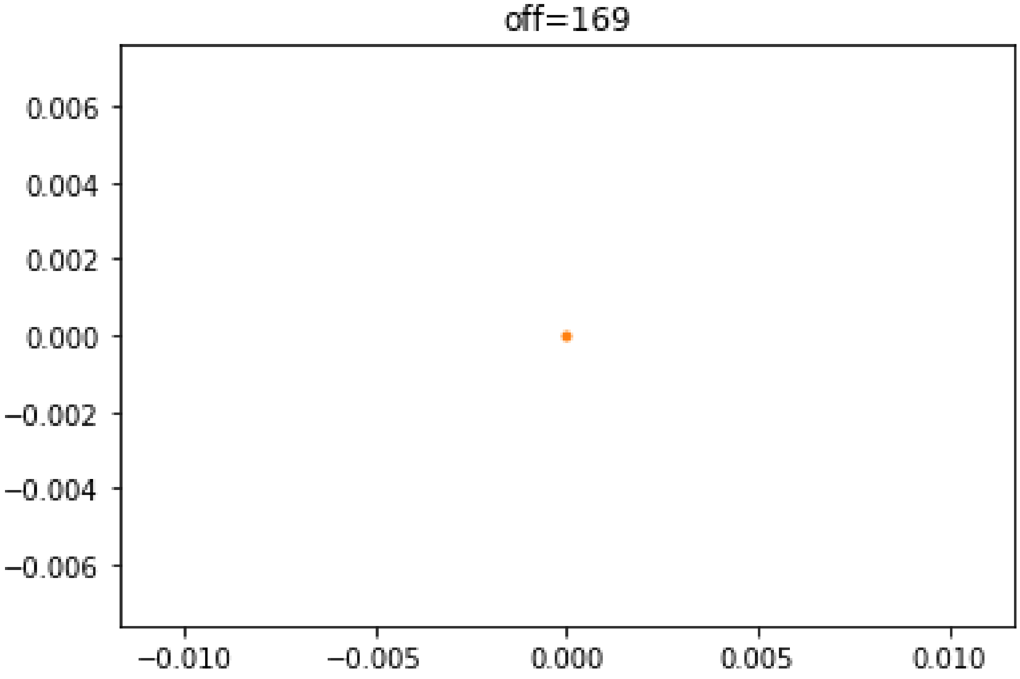

off=164
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

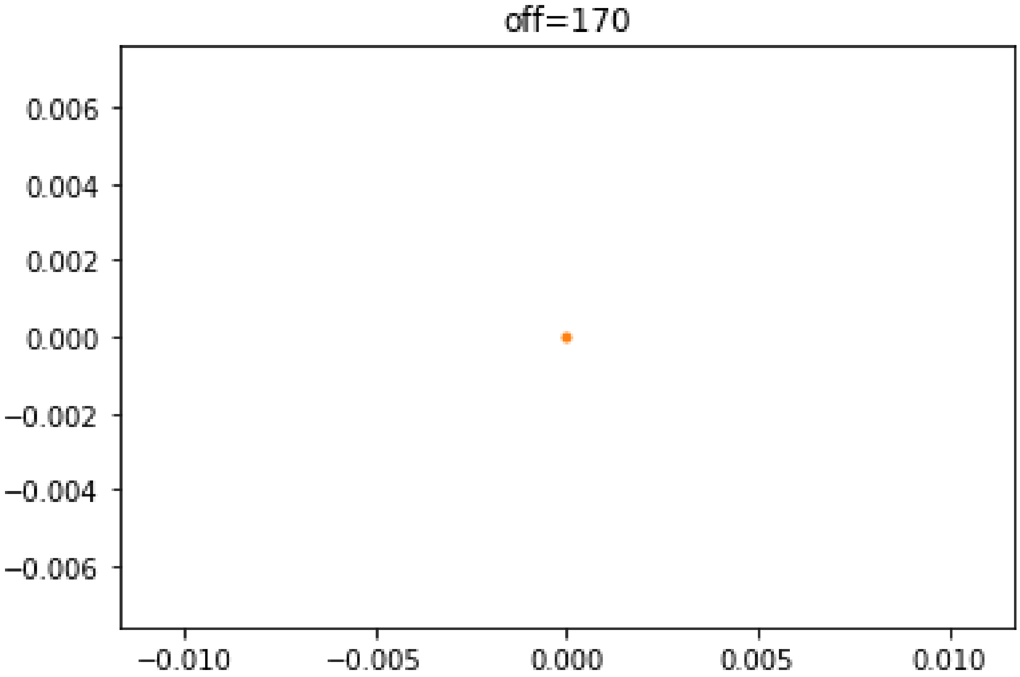

off=165
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

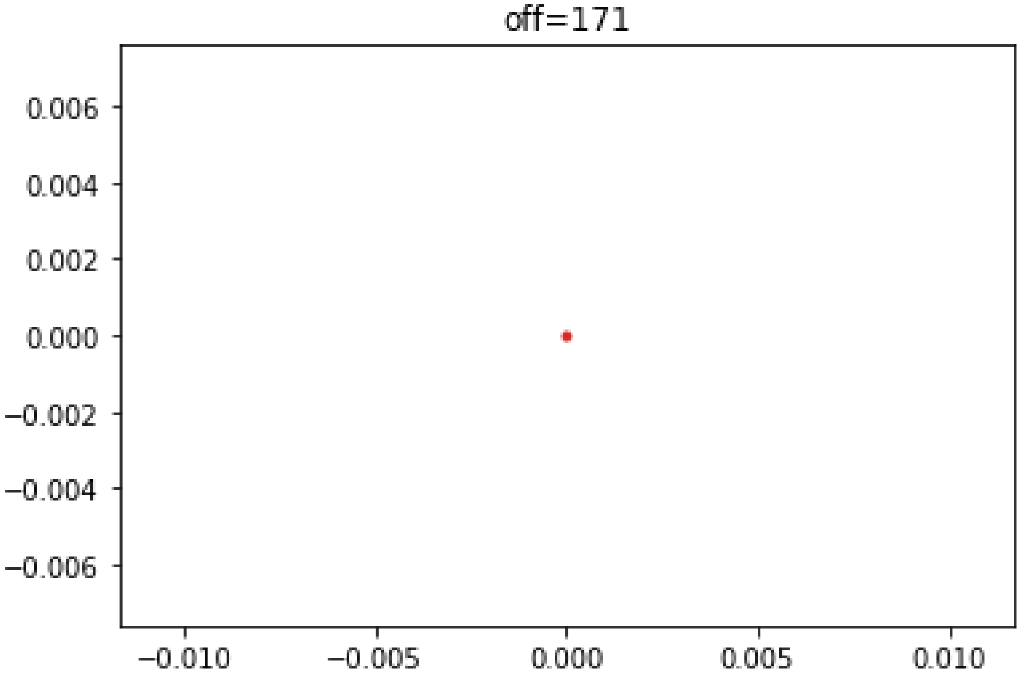

off=166
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

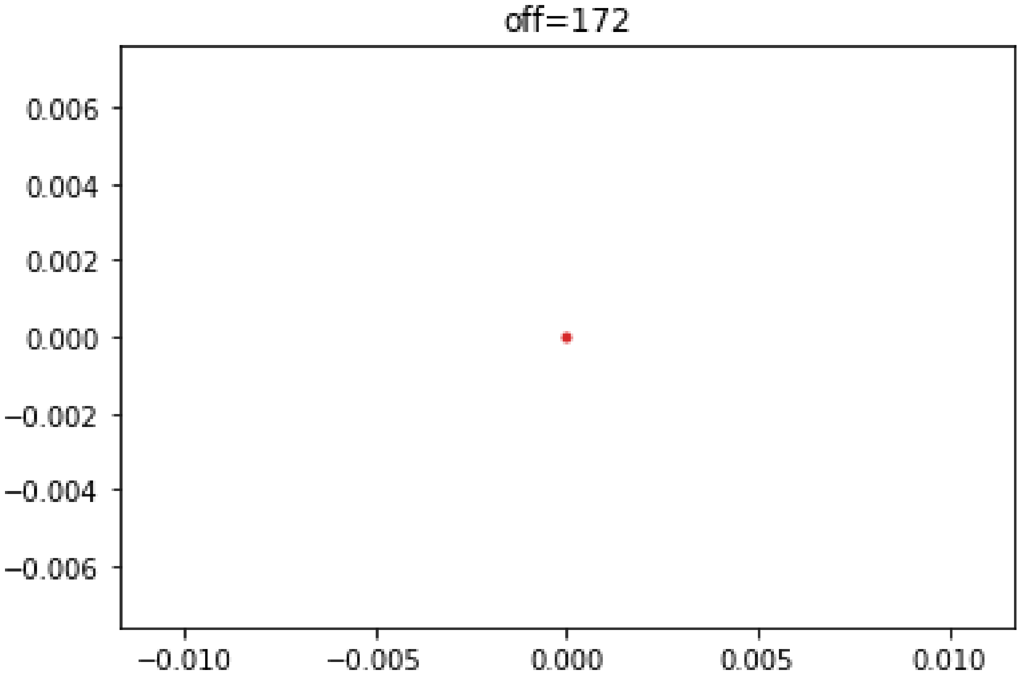

off=167
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

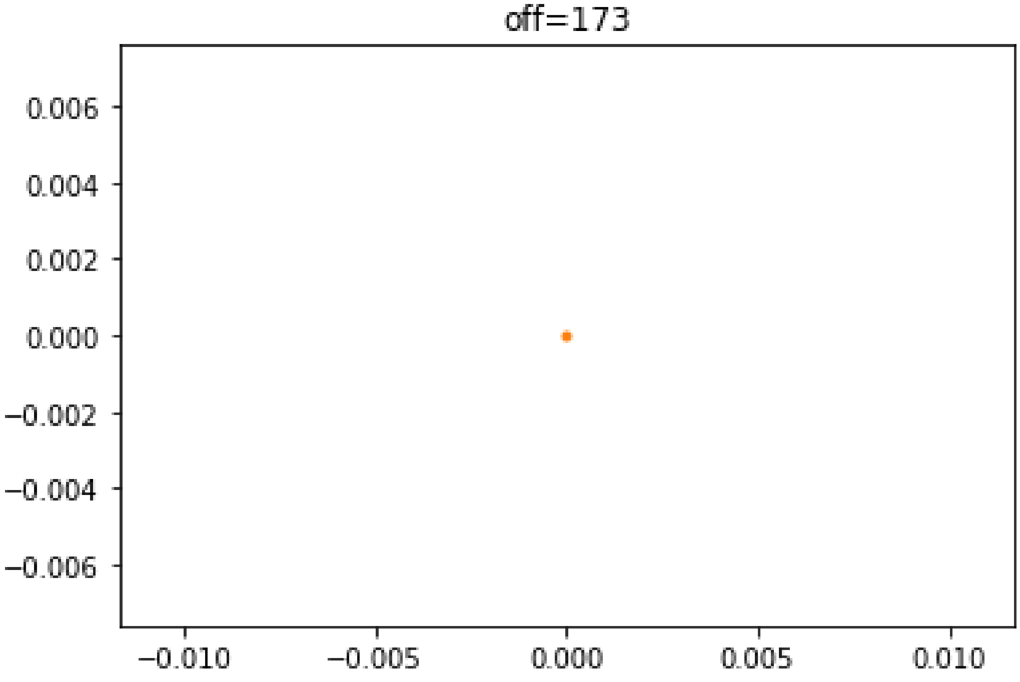

off=168
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

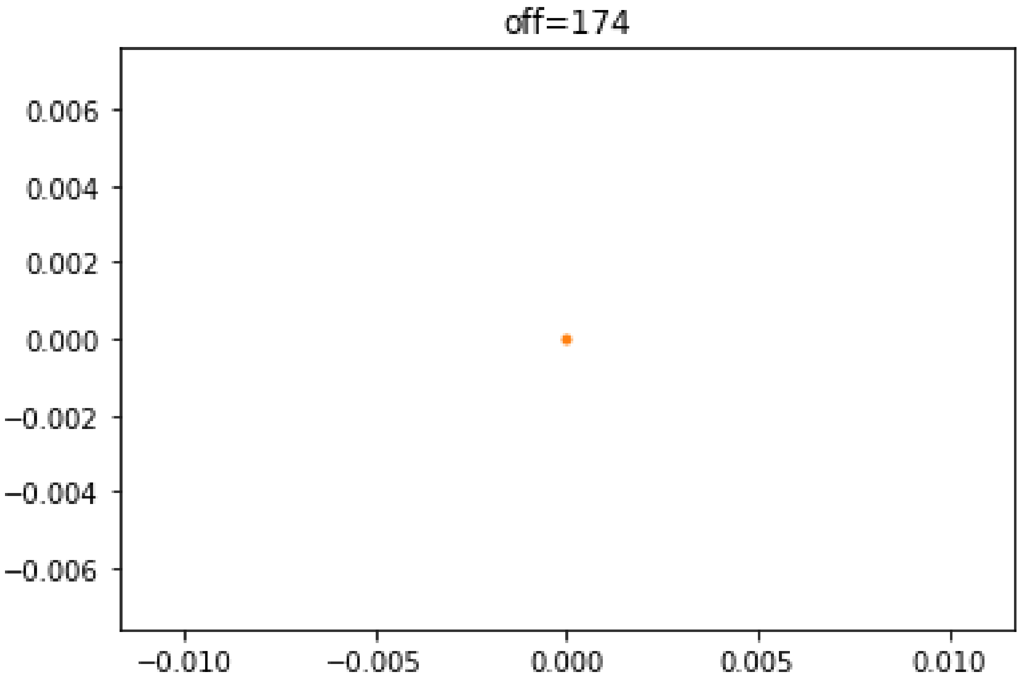

off=169
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010


off=170
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

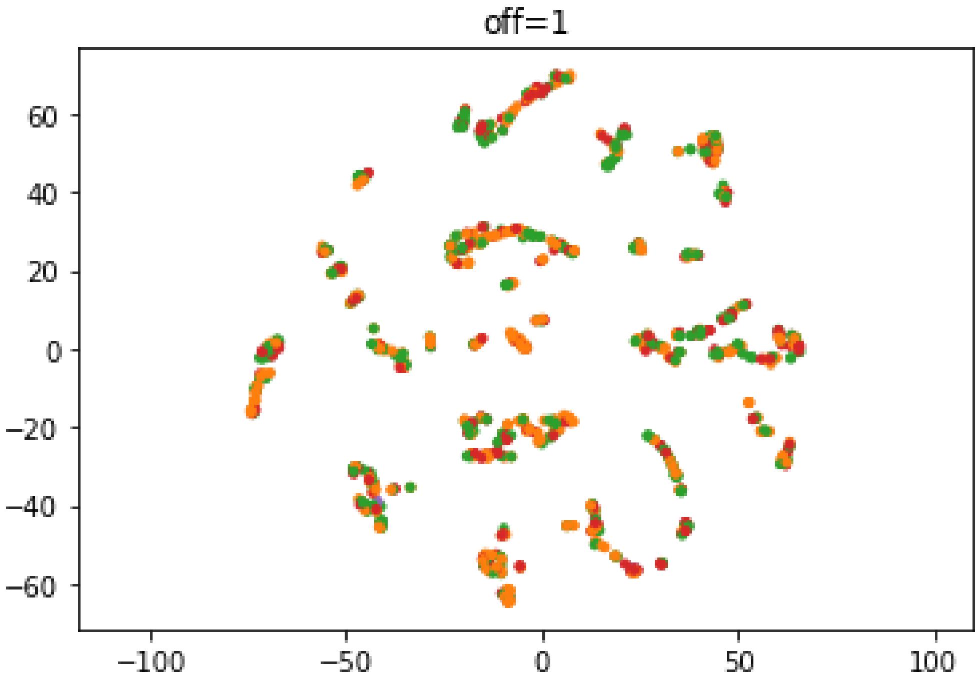

off=171
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

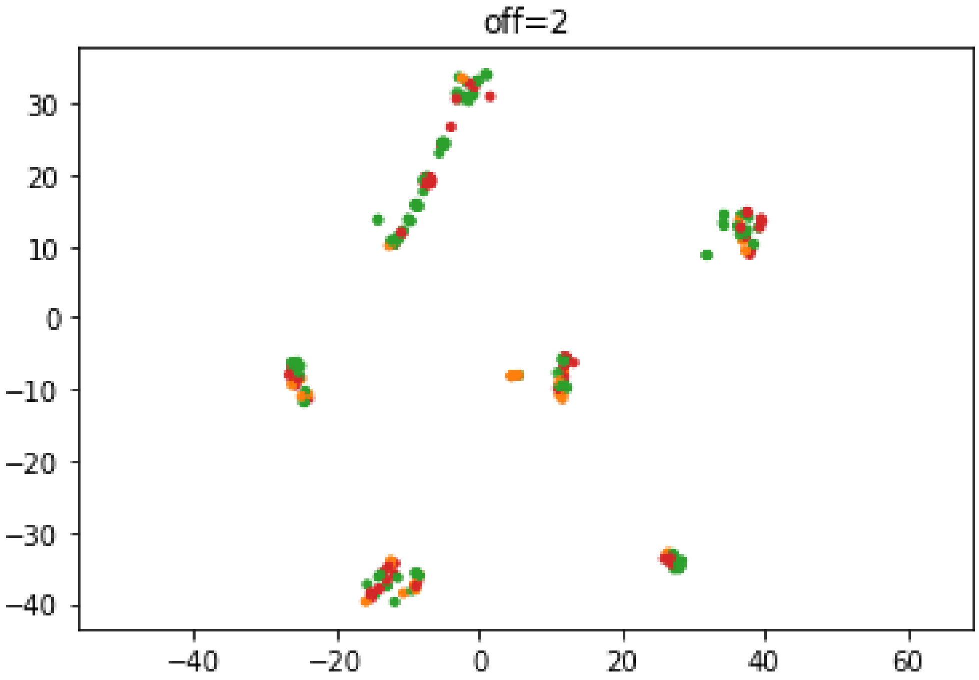

off=172
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

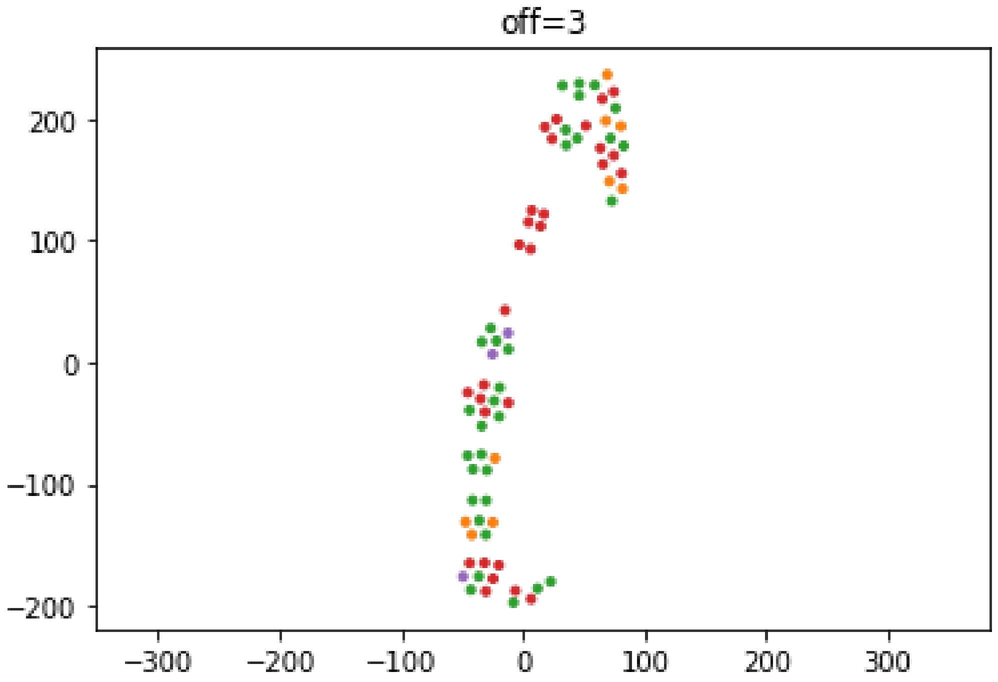

off=173
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.010
−0.005
0.000
0.005
0.010

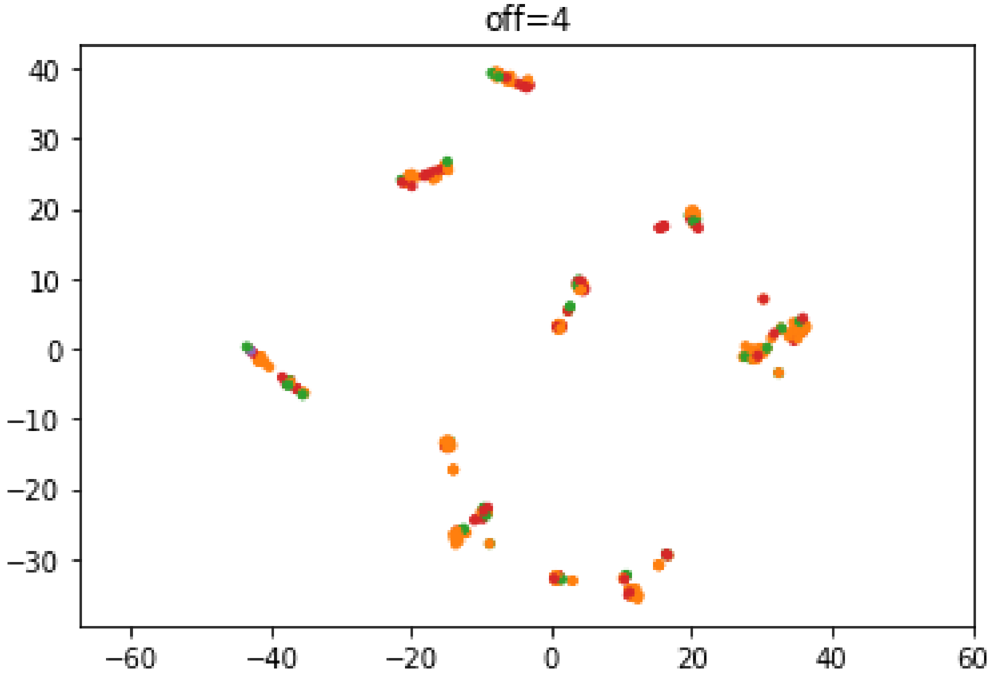


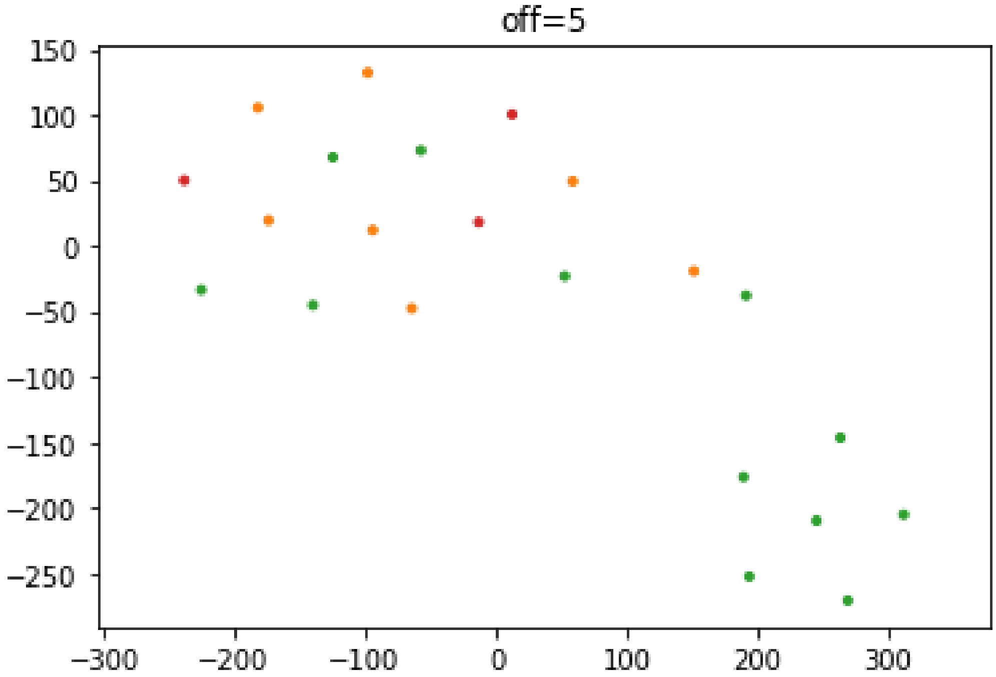
```
In [73]: for off in set(samp.OFFENSE):

             mask = samp_X.OFFENSE == off

             data = samp_X[mask]

             try:
                 trans = TSNE(perplexity=12).fit_transform(data)

                 plt.scatter(trans[:, 0], trans[:, 1], c=[sns.color_palette()[x] for x in samp_
                 plt.gca().set_aspect('equal', 'datalim')
                 plt.title(f'off={off}')
                 plt.show()
             except:

                 continue
```

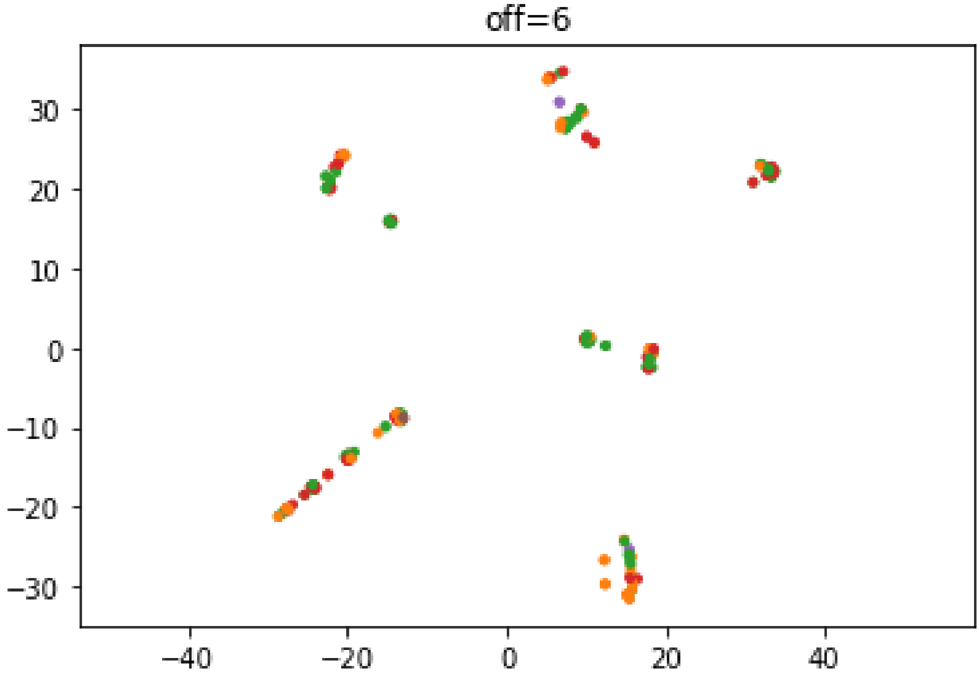

off=1
60
40
20
0
−20
−40
−60
−100
−50
0
50
100

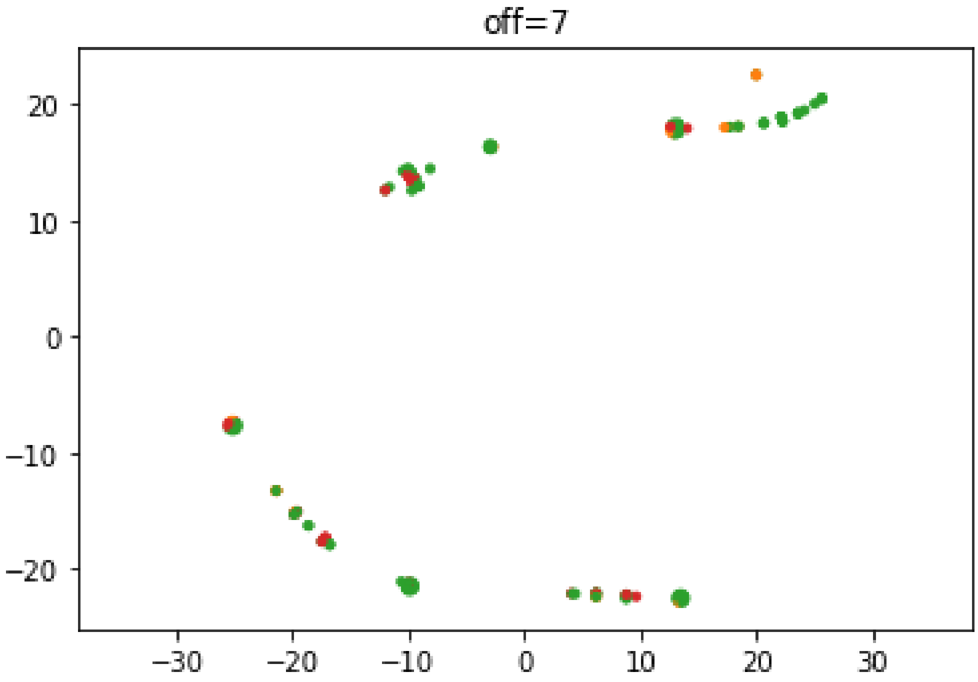

off=2
30
20
10
0
−10
−20
−30
−40
−40
−20
0
20
40
60

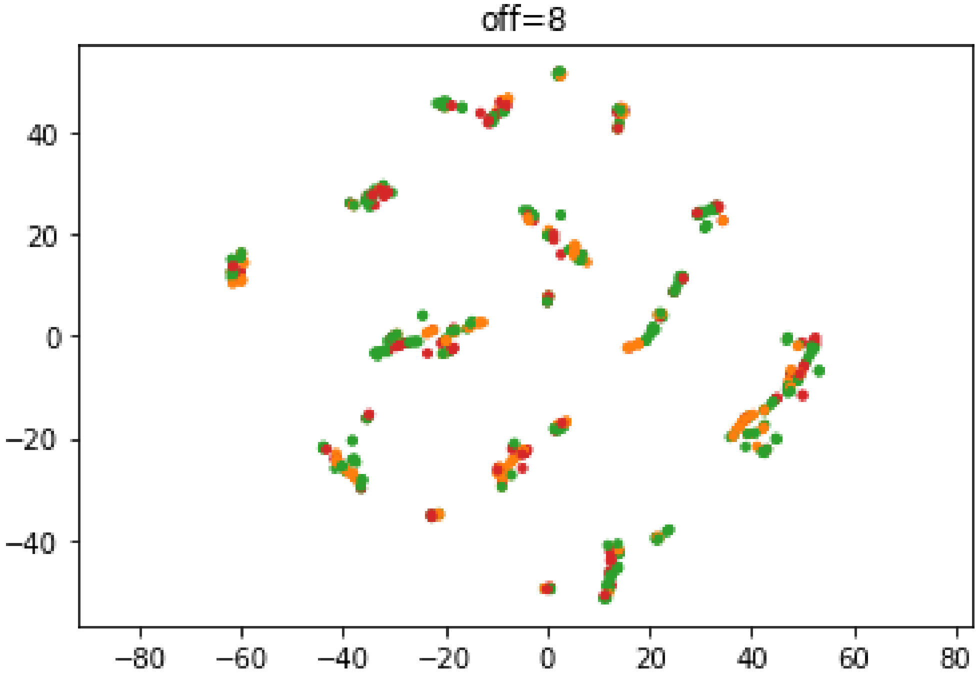

off=3
200
100
0
−100
−200
−300
−200
−100
0
100
200
300

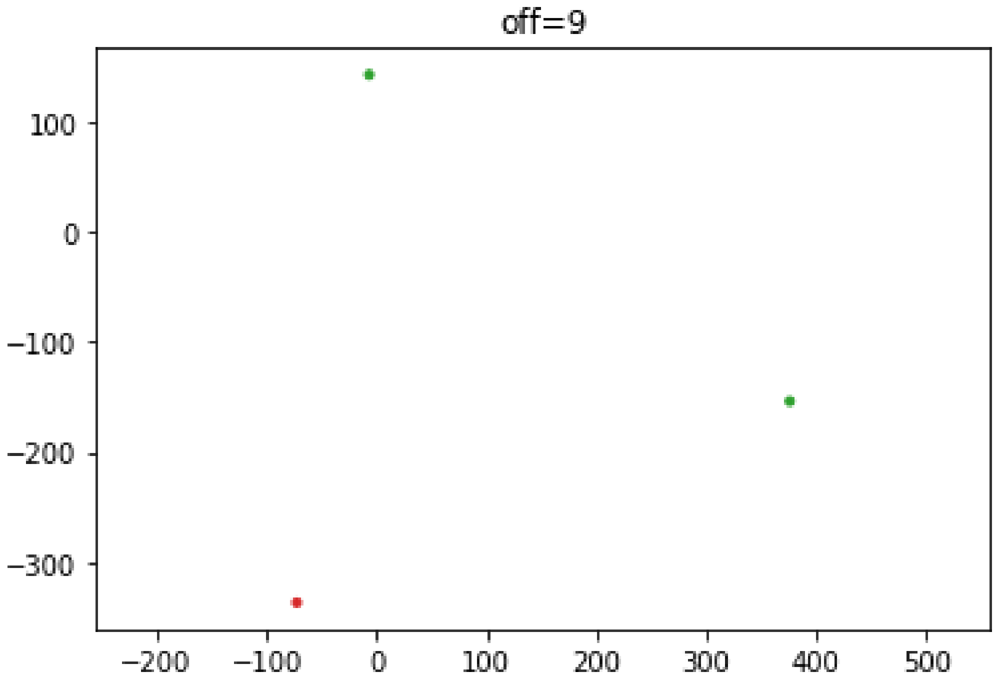

off=4
40
30
20
10
0
−10
−20
−30
−60
−40
−20
0
20
40
60

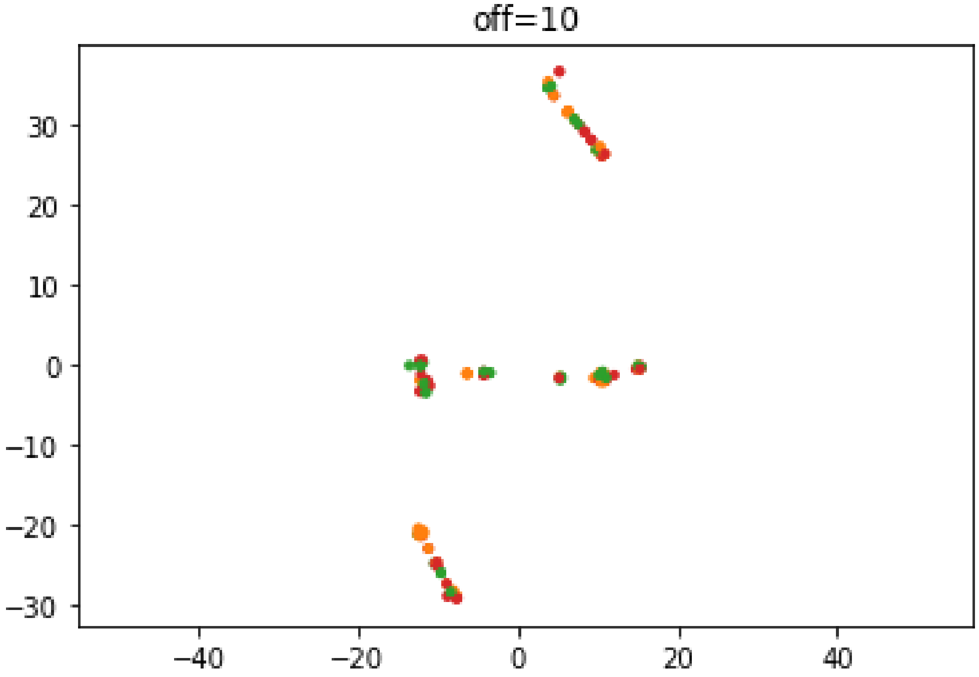

off=5
150
100
50
0
−50
−100
−150
−200
−250
−300
−200
−100
0
100
200
300

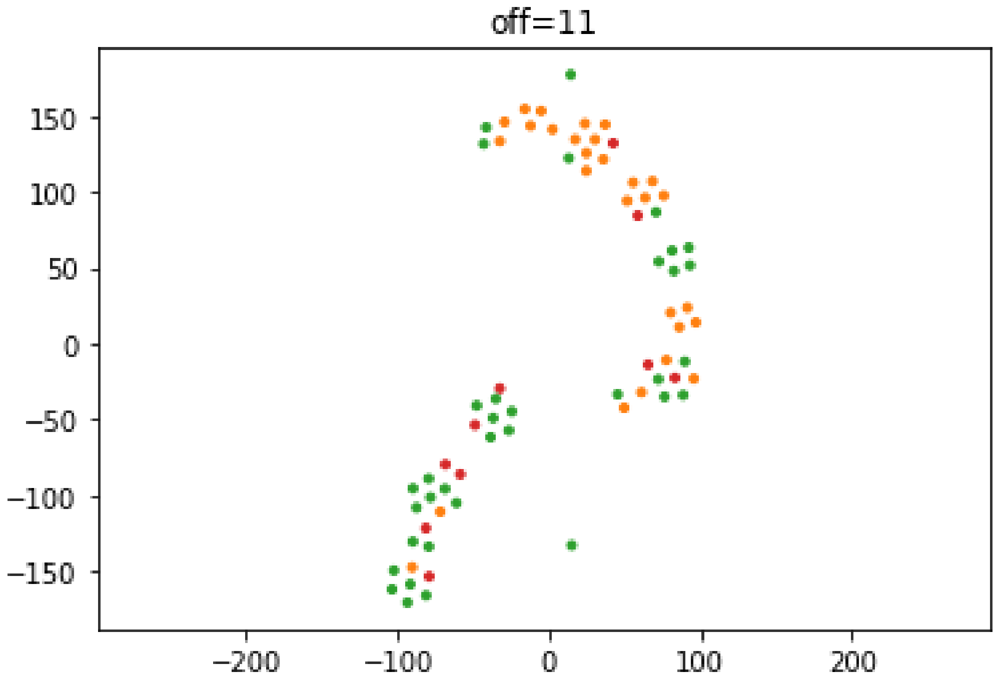

off=6
30
20
10
0
−10
−20
−30
−40
−20
0
20
40

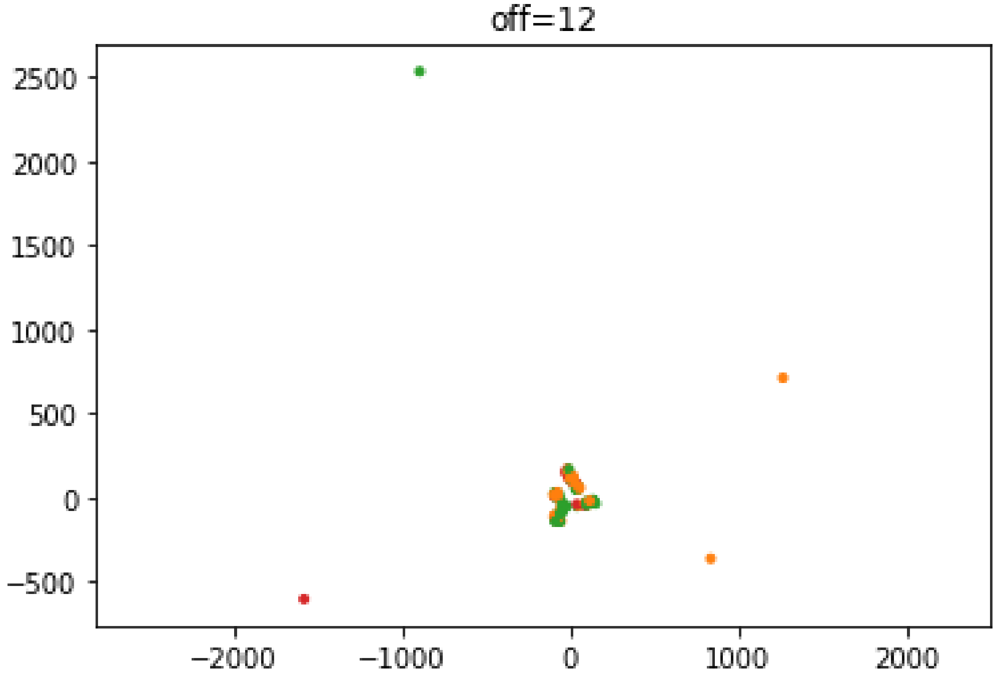

off=7
20
10
0
−10
−20
−30
−20
−10
0
10
20
30

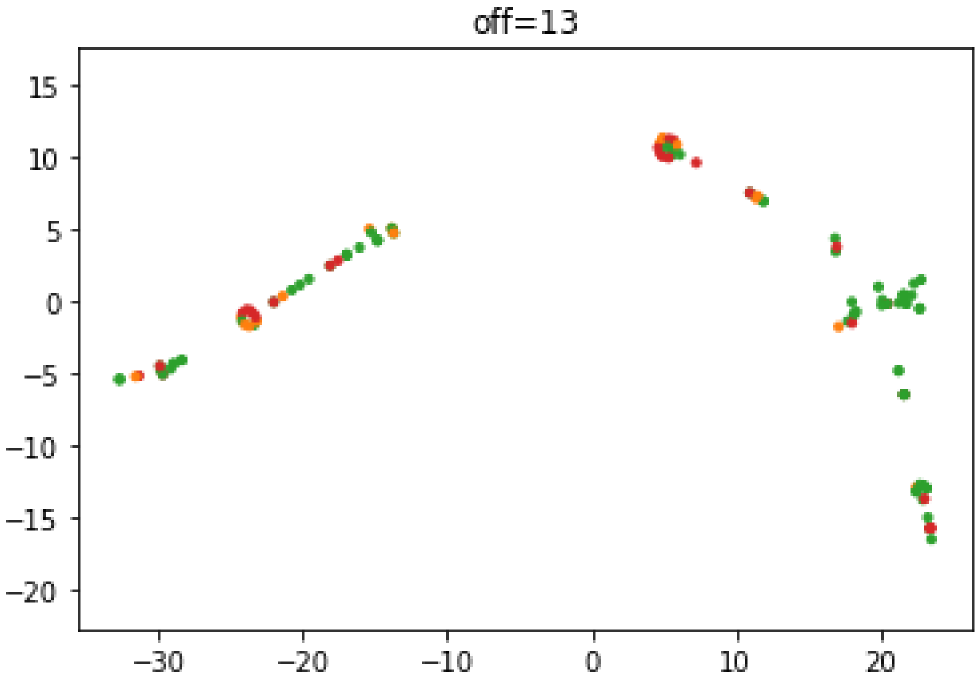

off=8
40
20
0
−20
−40
−80
−60
−40
−20
0
20
40
60
80

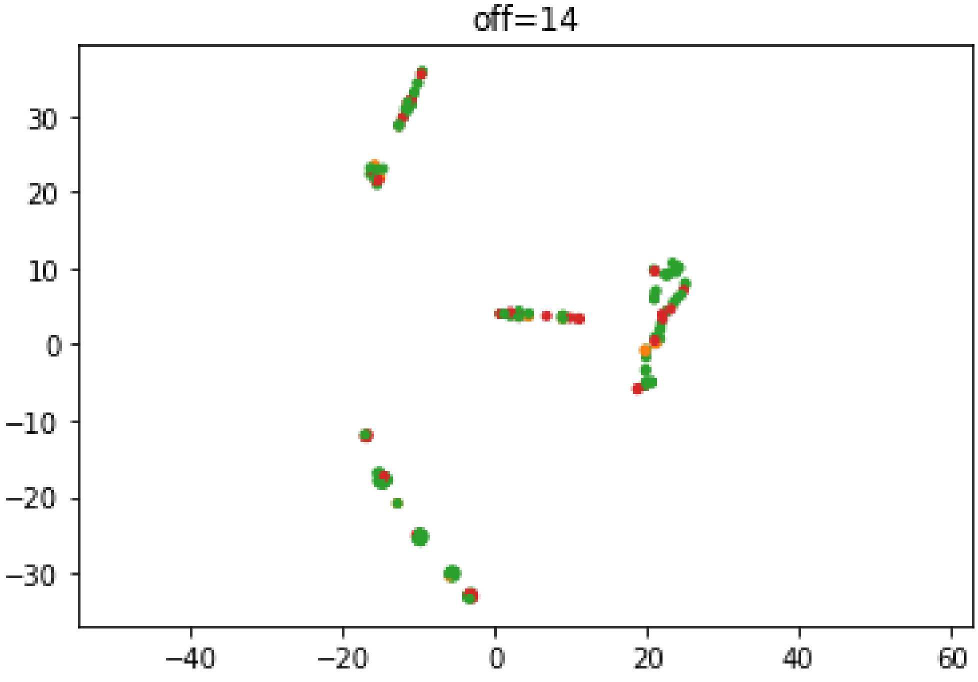

off=9
100
0
−100
−200
−300
−200
−100
0
100
200
300
400
500

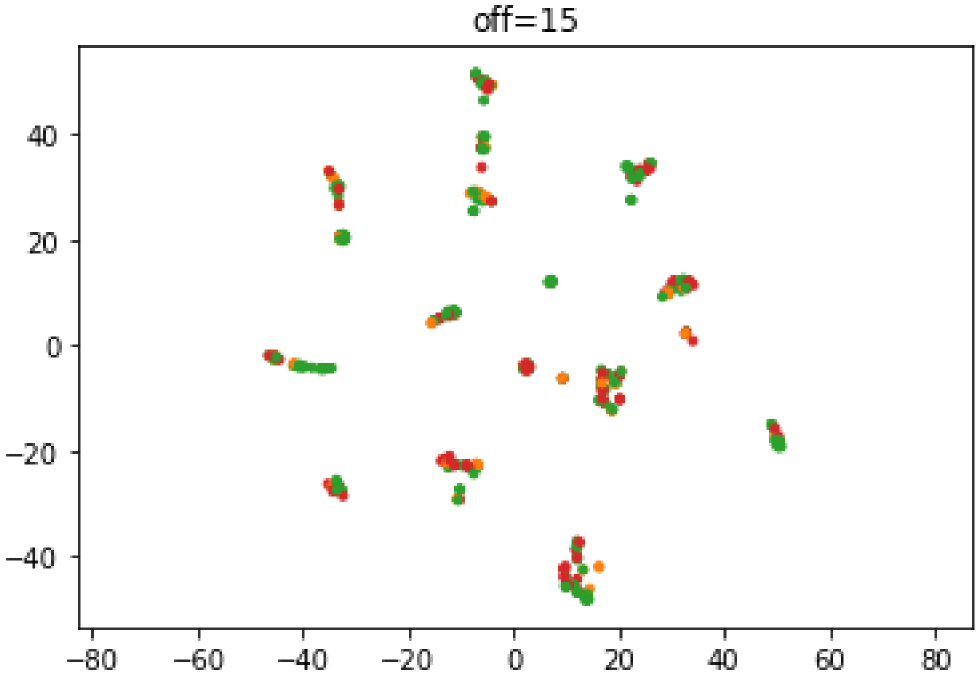

off=10
30
20
10
0
−10
−20
−30
−40
−20
0
20
40

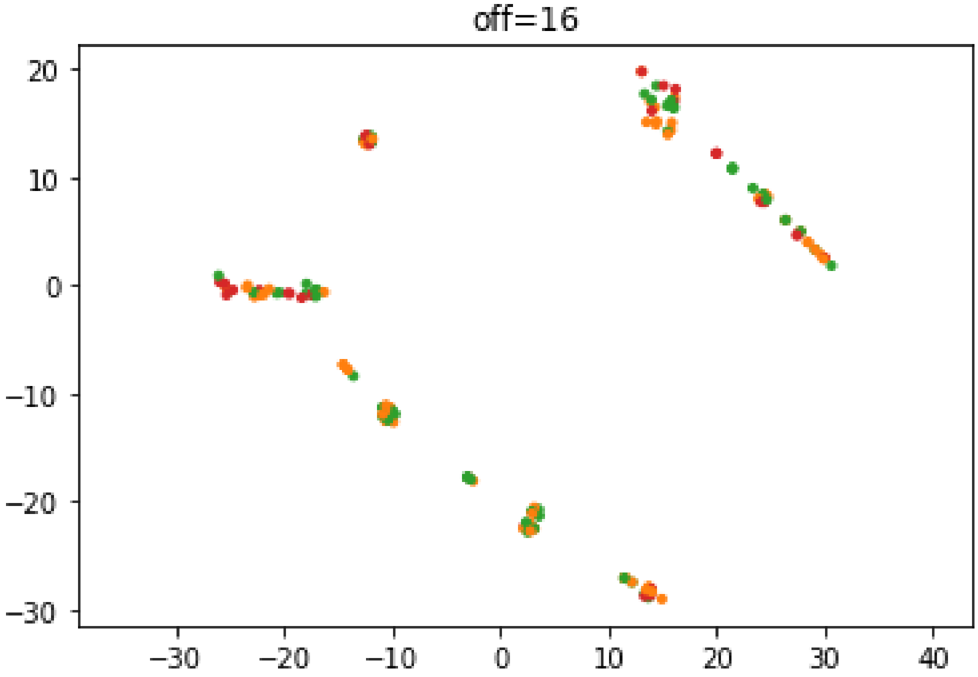

off=11
150
100
50
0
−50
−100
−150
−200
−100
0
100
200

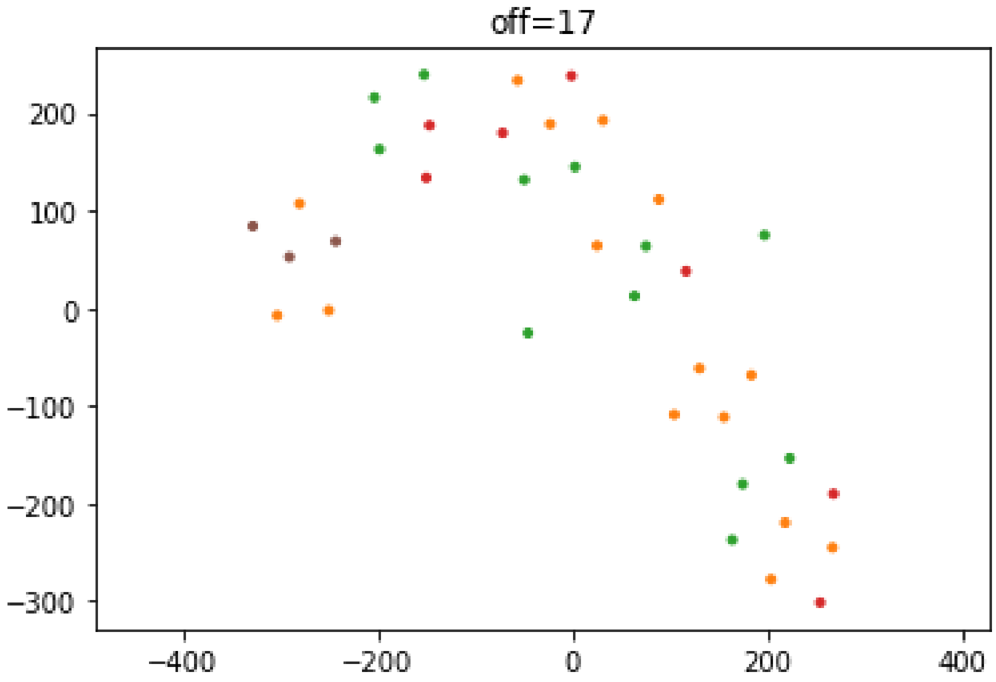

off=12
2500
2000
1500
1000
500
0
−500
−2000
−1000
0
1000
2000

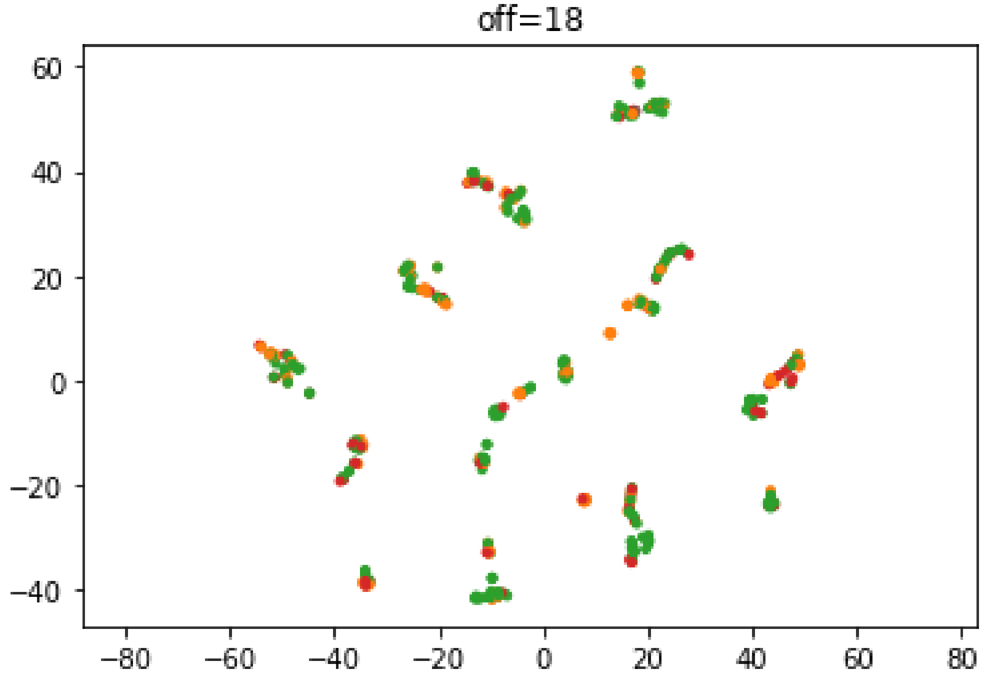

off=13
15
10
5
0
−5
−10
−15
−20
−30
−20
−10
0
10
20

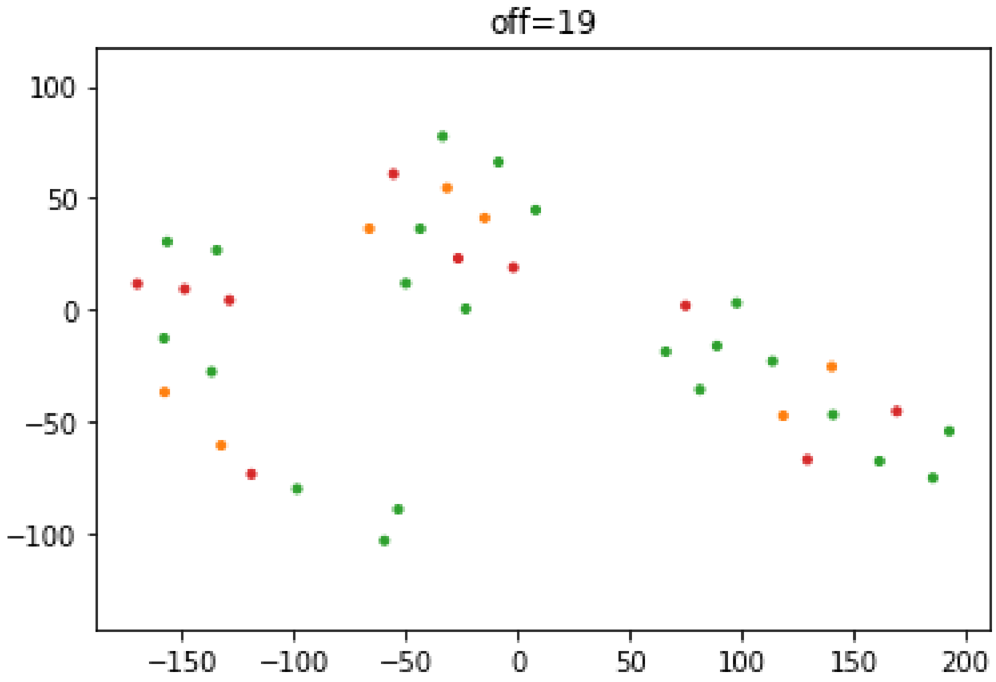

off=14
30
20
10
0
−10
−20
−30
−40
−20
0
20
40
60

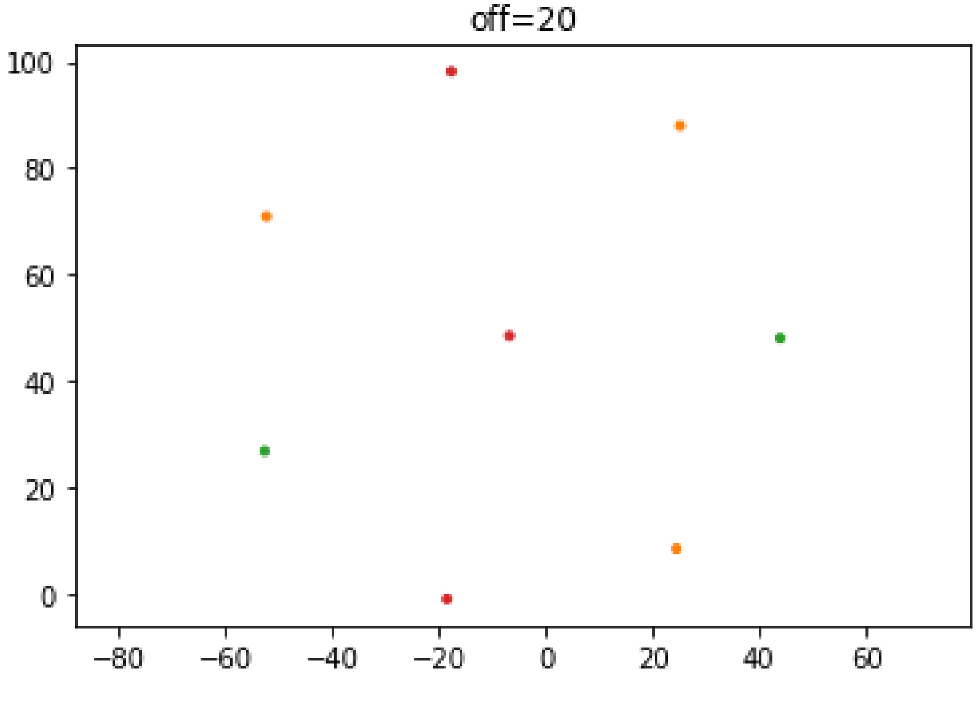

off=15
40
20
0
−20
−40
−80
−60
−40
−20
0
20
40
60
80

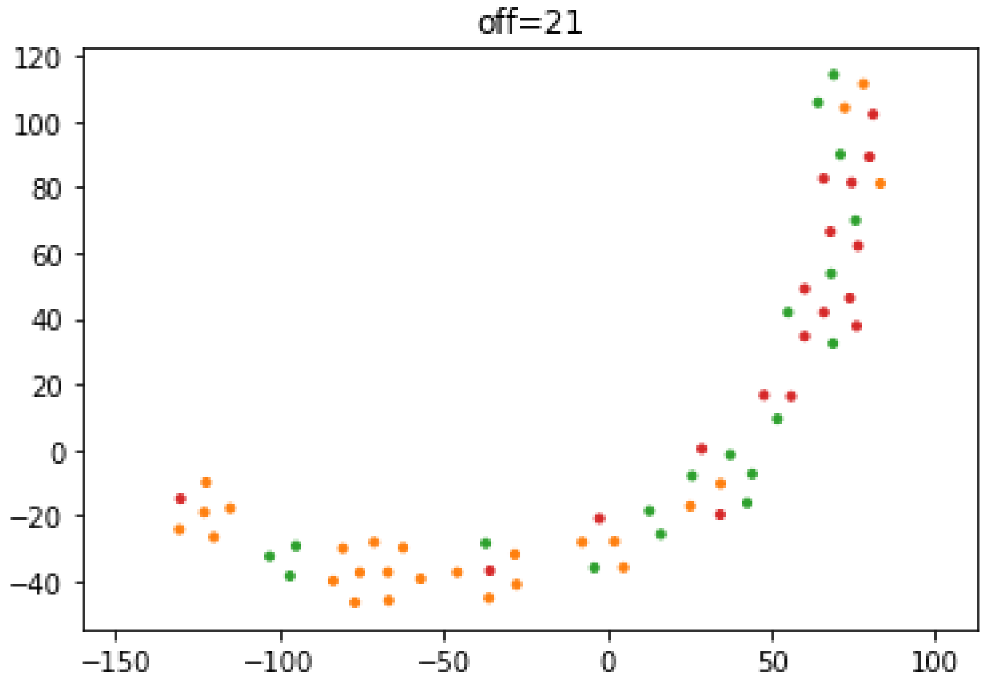

off=16
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30
40

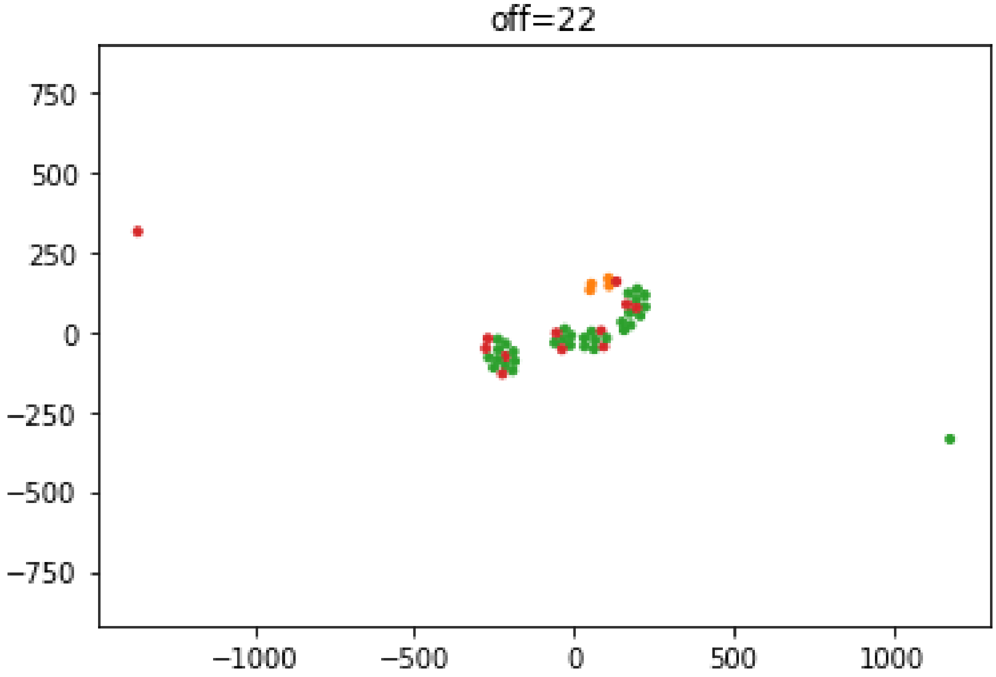

off=17
200
100
0
−100
−200
−300
−400
−200
0
200
400

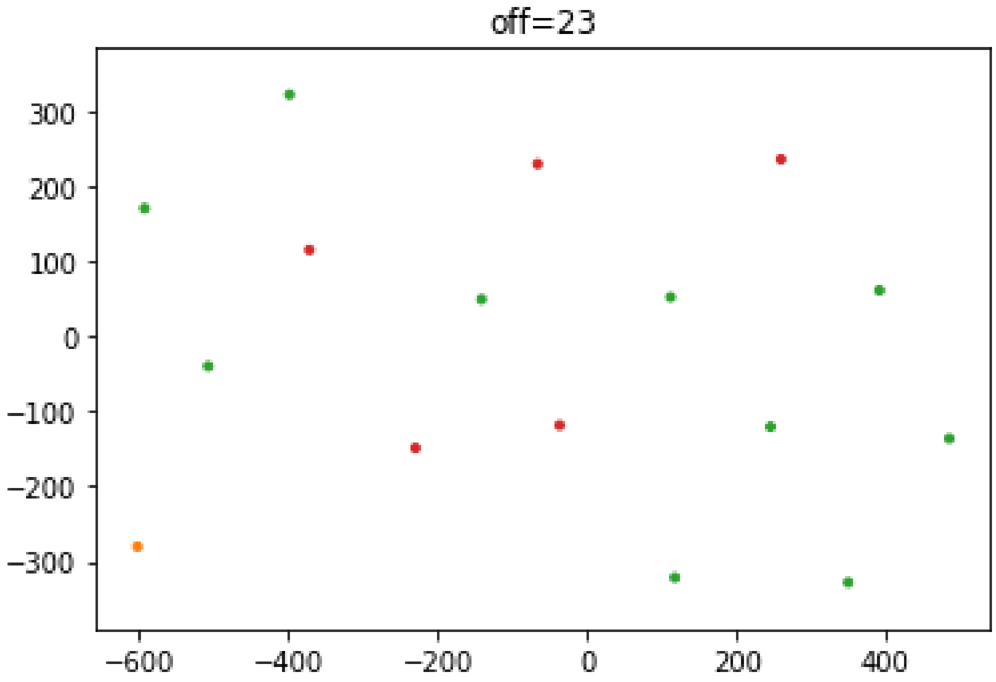

off=18
60
40
20
0
−20
−40
−80
−60
−40
−20
0
20
40
60
80

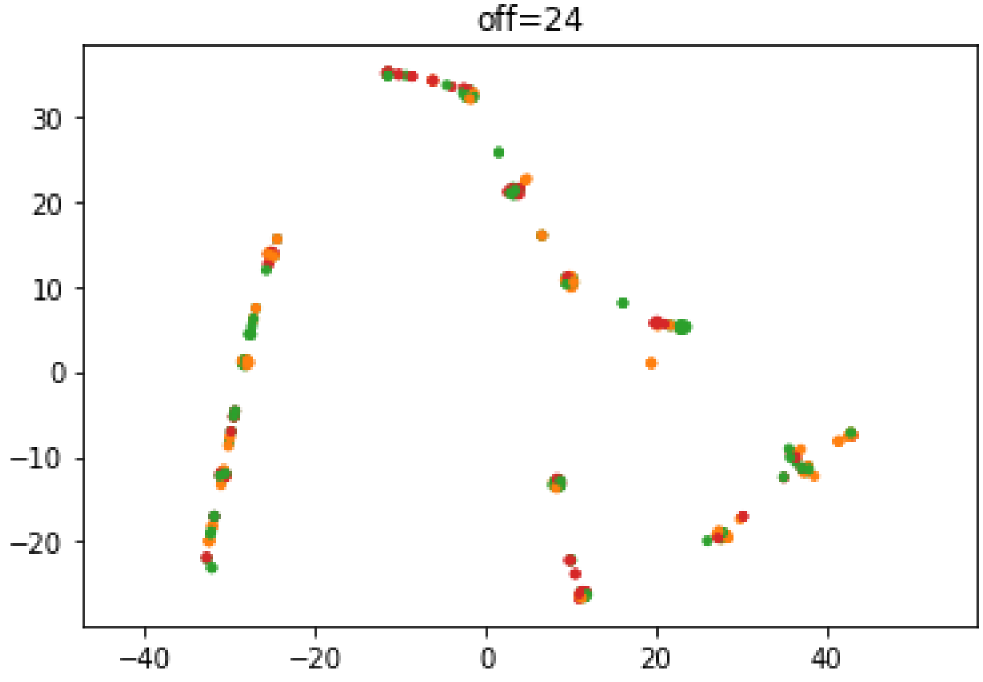

off=19
100
50
0
−50
−100
−150
−100
−50
0
50
100
150
200

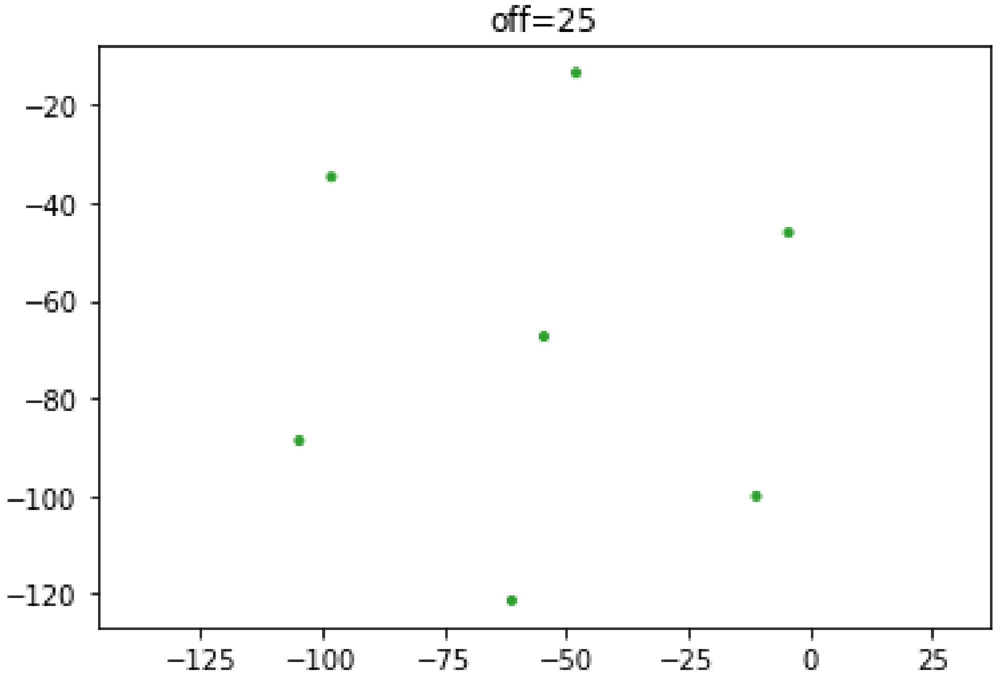

off=20
100
80
60
40
20
0
−80
−60
−40
−20
0
20
40
60

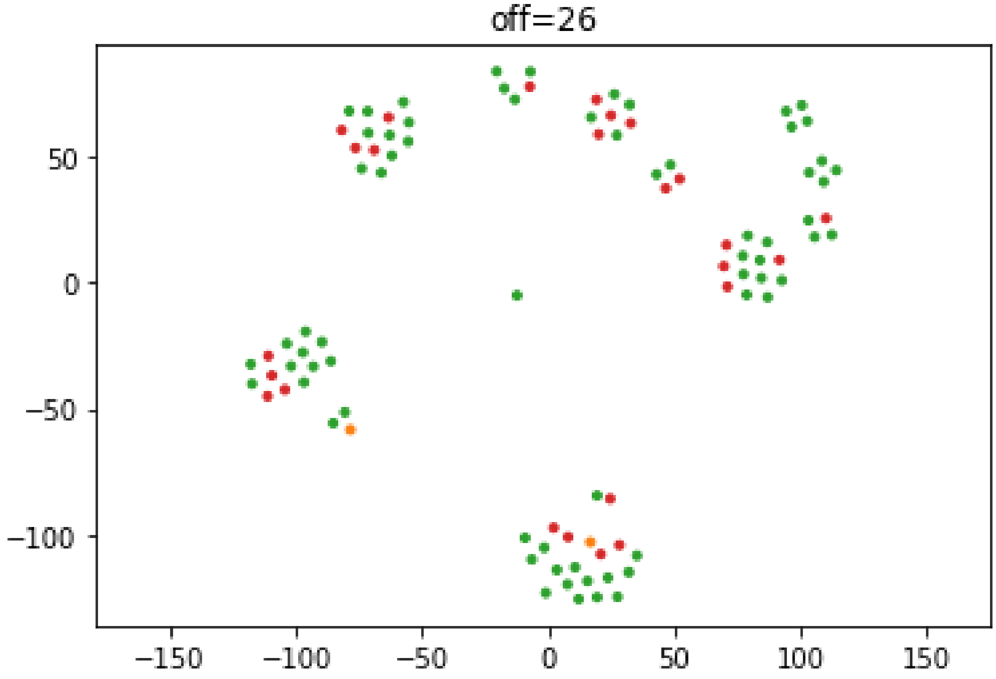

off=21
120
100
80
60
40
20
0
−20
−40
−150
−100
−50
0
50
100

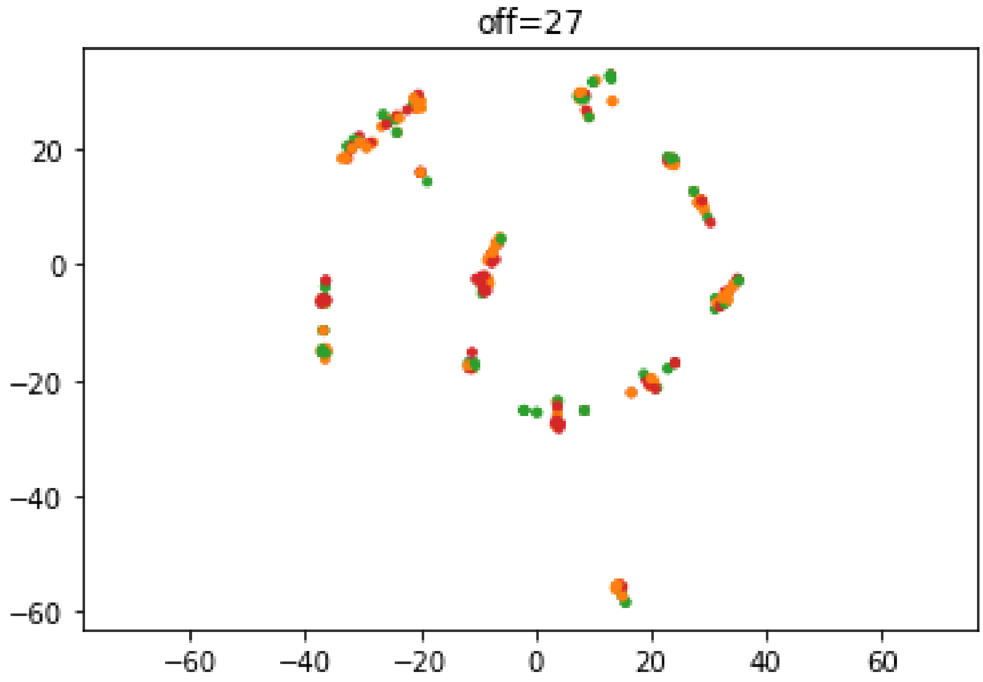

off=22
750
500
250
0
−250
−500
−750
−1000
−500
0
500
1000

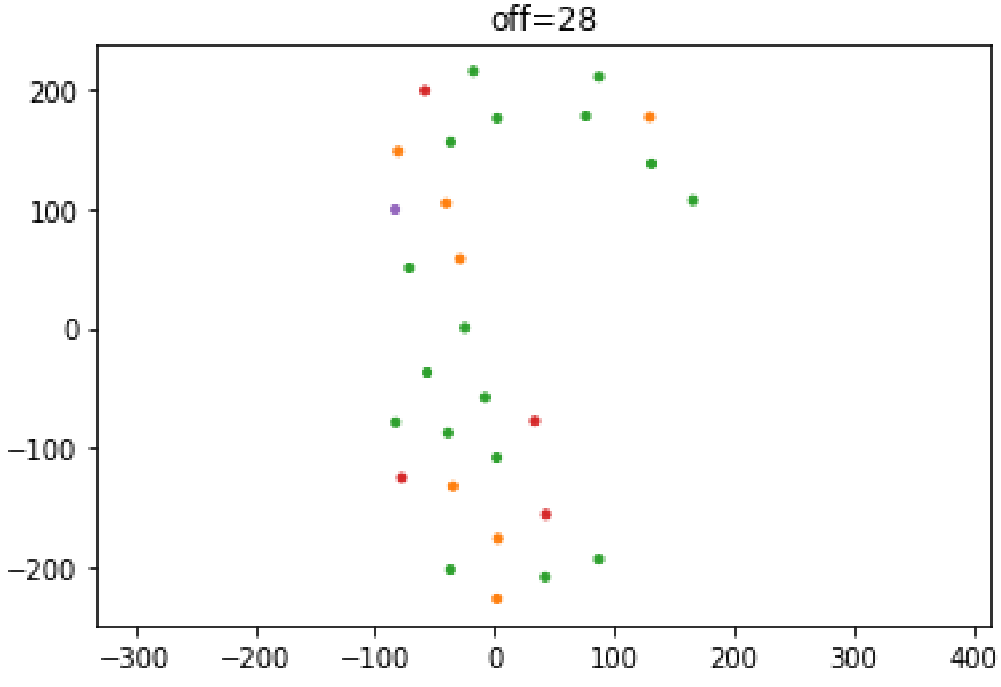

off=23
300
200
100
0
−100
−200
−300
−600
−400
−200
0
200
400

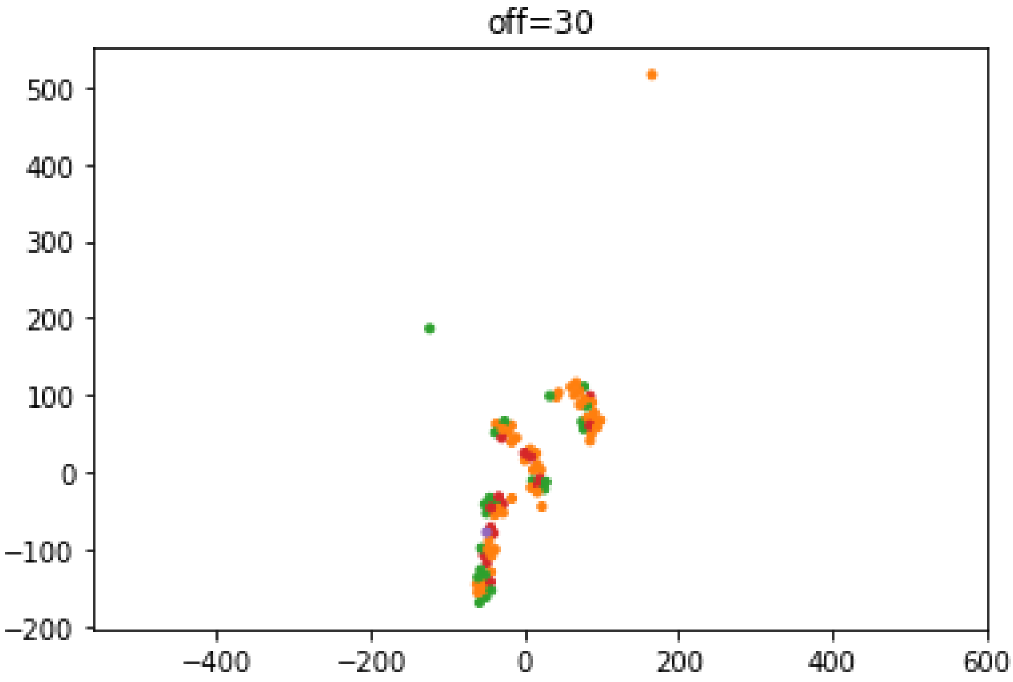

off=24
30
20
10
0
−10
−20
−40
−20
0
20
40

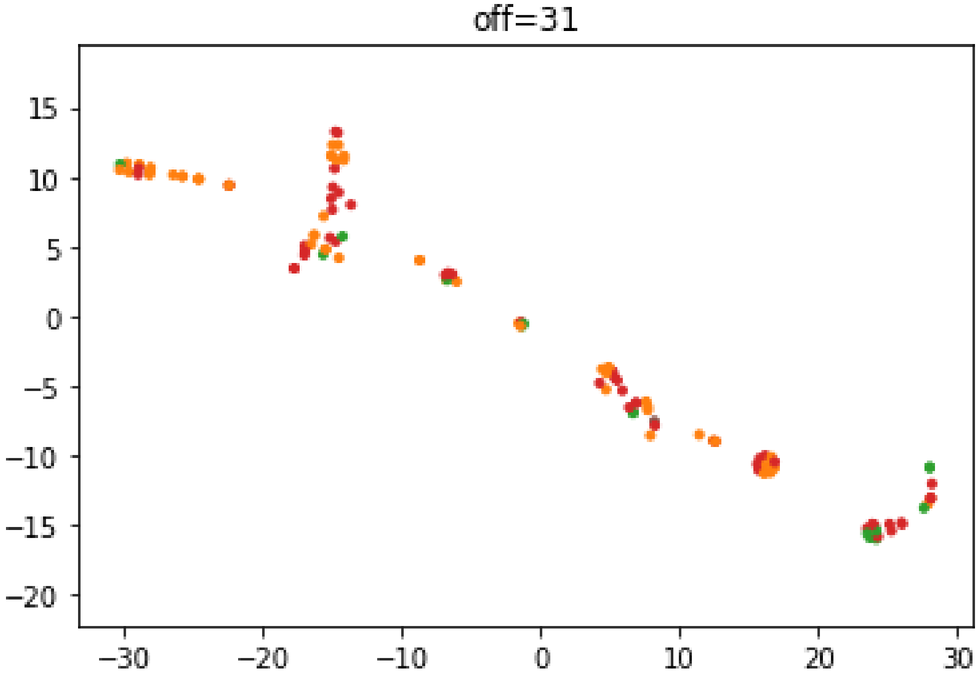

off=25
−20
−40
−60
−80
−100
−120
−125
−100
−75
−50
−25
0
25

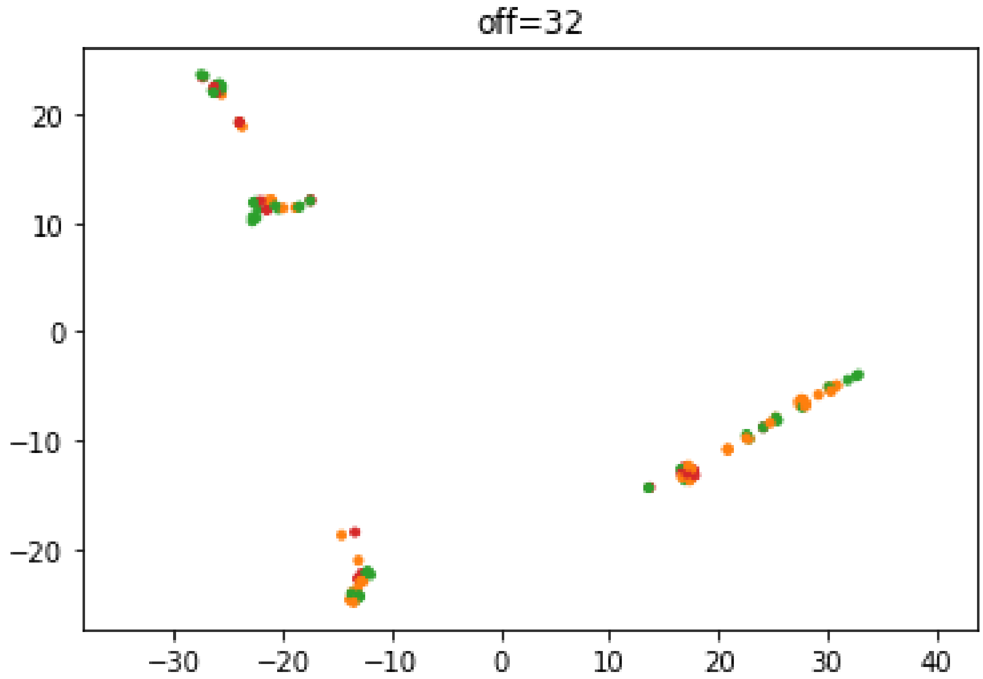

off=26
50
0
−50
−100
−150
−100
−50
0
50
100
150

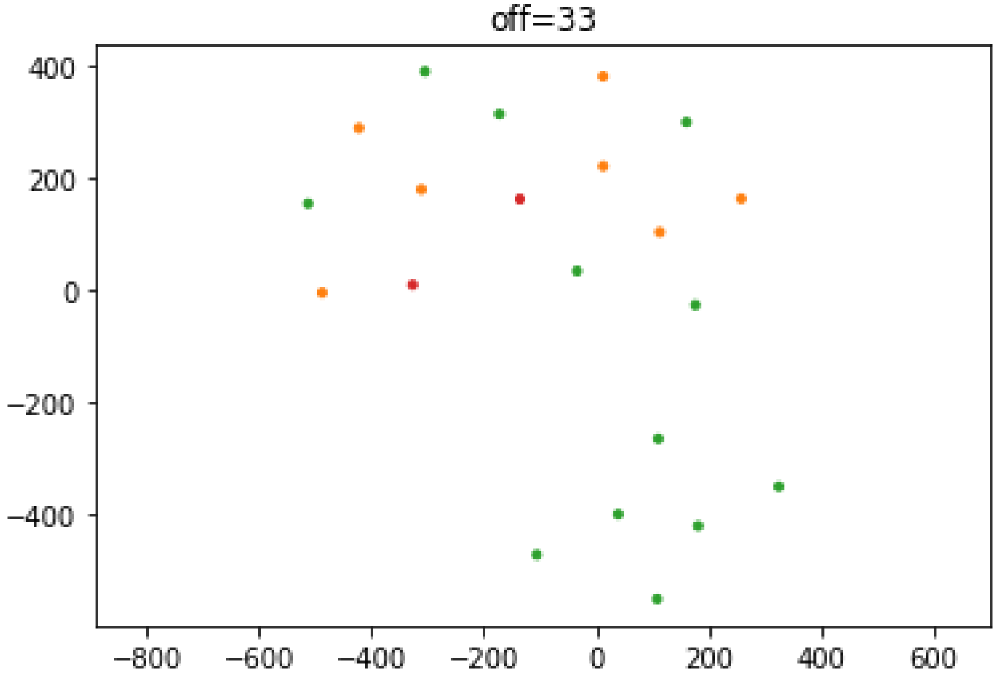

off=27
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60

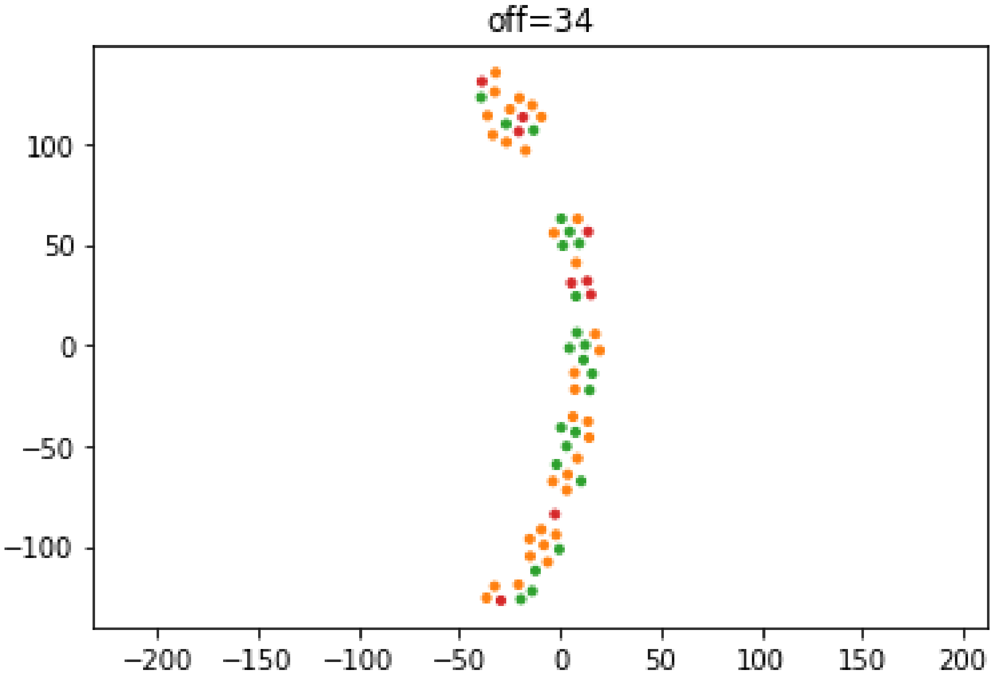

off=28
200
100
0
−100
−200
−300
−200
−100
0
100
200
300
400

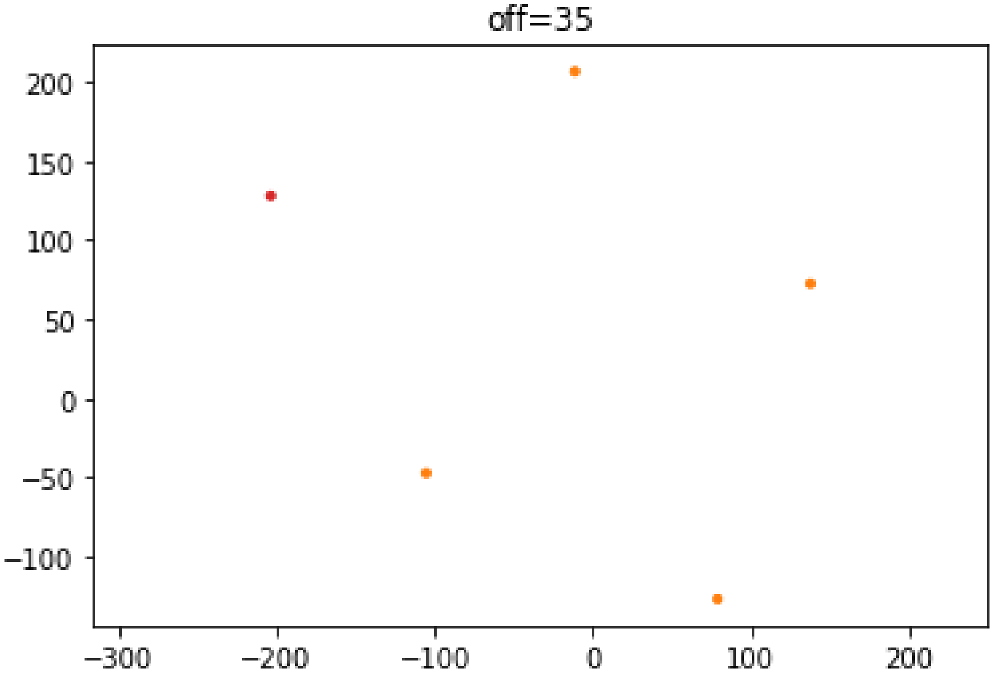

off=30
500
400
300
200
100
0
−100
−200
−400
−200
0
200
400
600

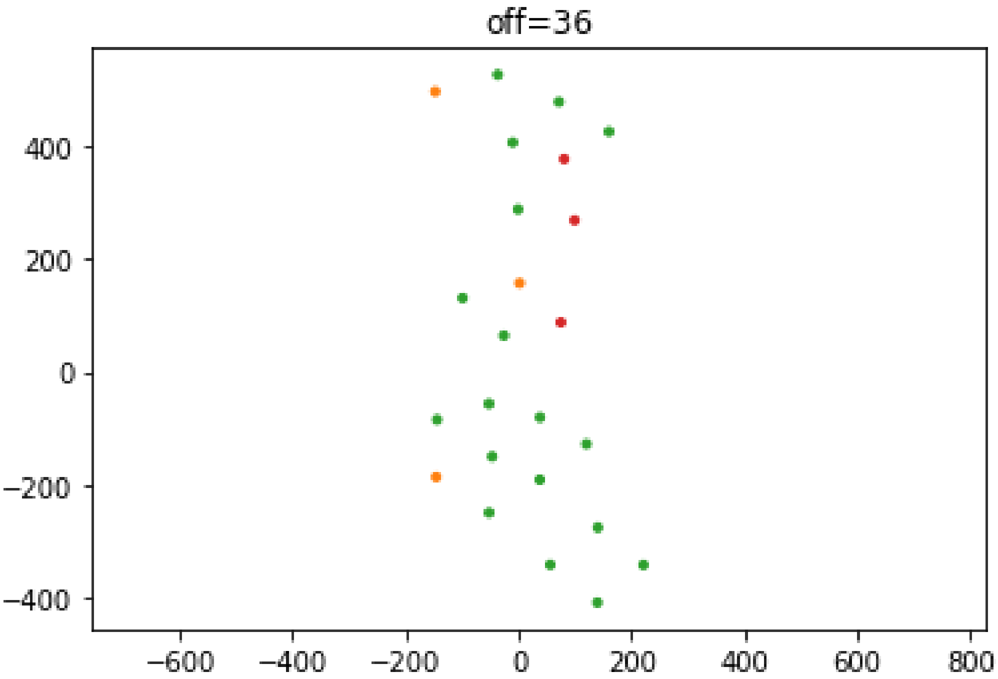

off=31
15
10
5
0
−5
−10
−15
−20
−30
−20
−10
0
10
20
30

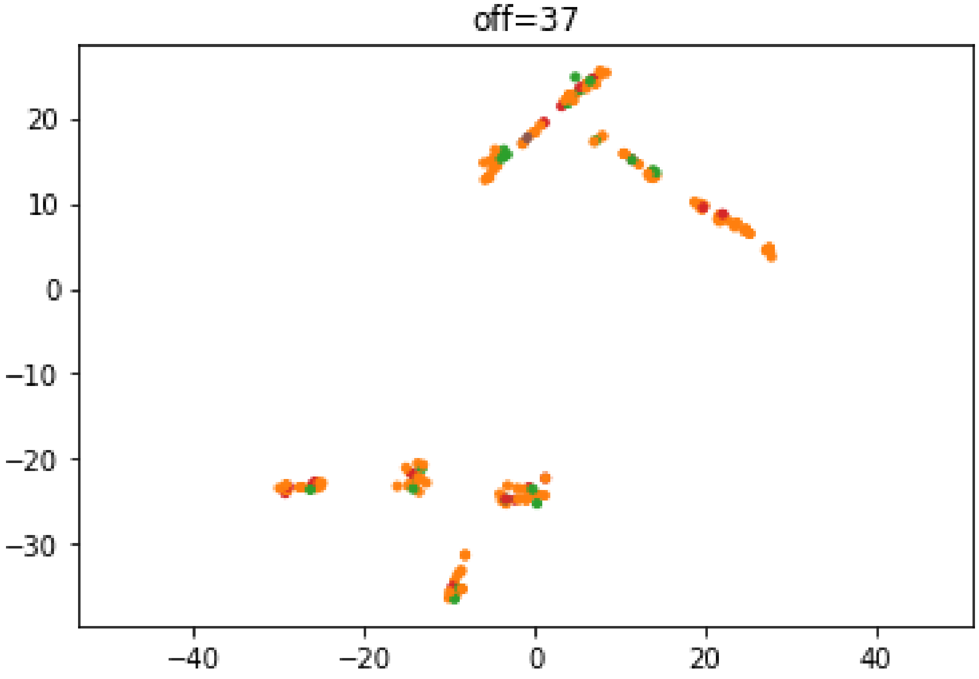

off=32
20
10
0
−10
−20
−30
−20
−10
0
10
20
30
40

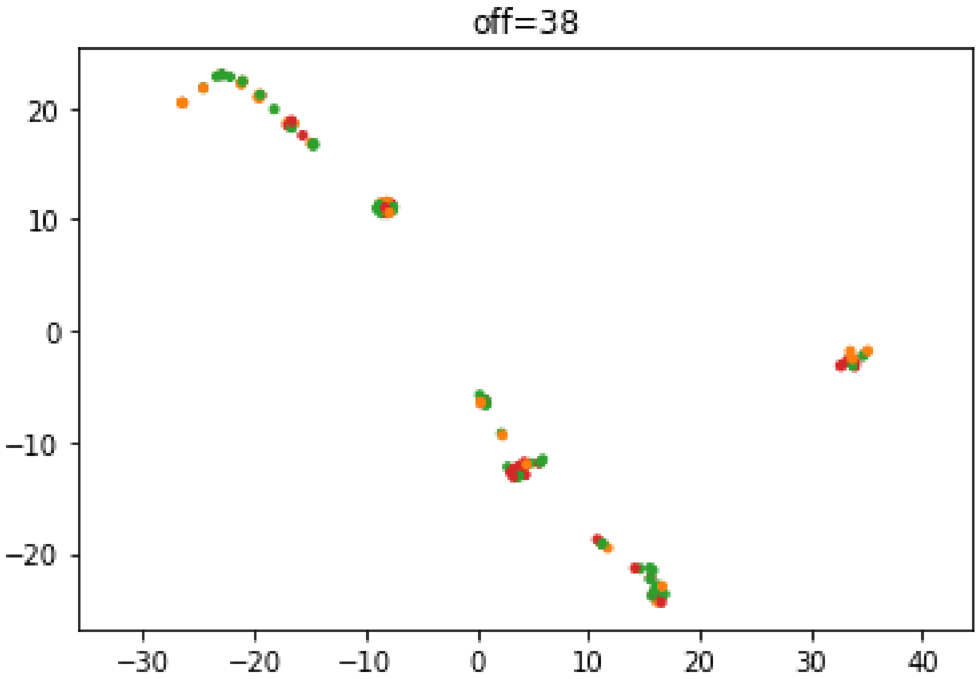

off=33
400
200
0
−200
−400
−800
−600
−400
−200
0
200
400
600

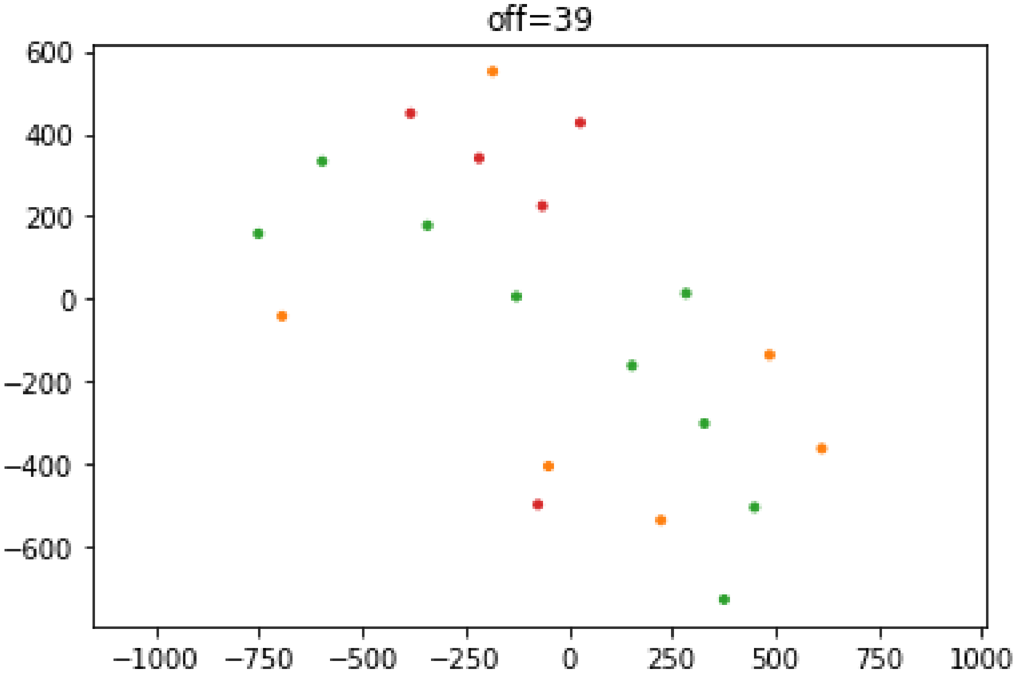

off=34
100
50
0
−50
−100
−200
−150
−100
−50
0
50
100
150
200

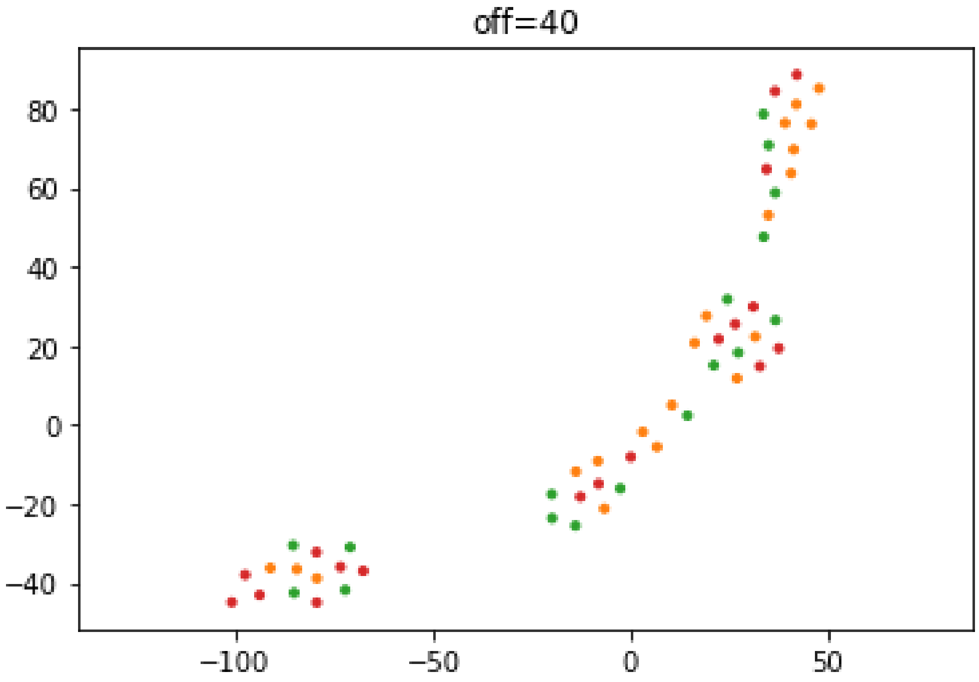

off=35
200
150
100
50
0
−50
−100
−300
−200
−100
0
100
200

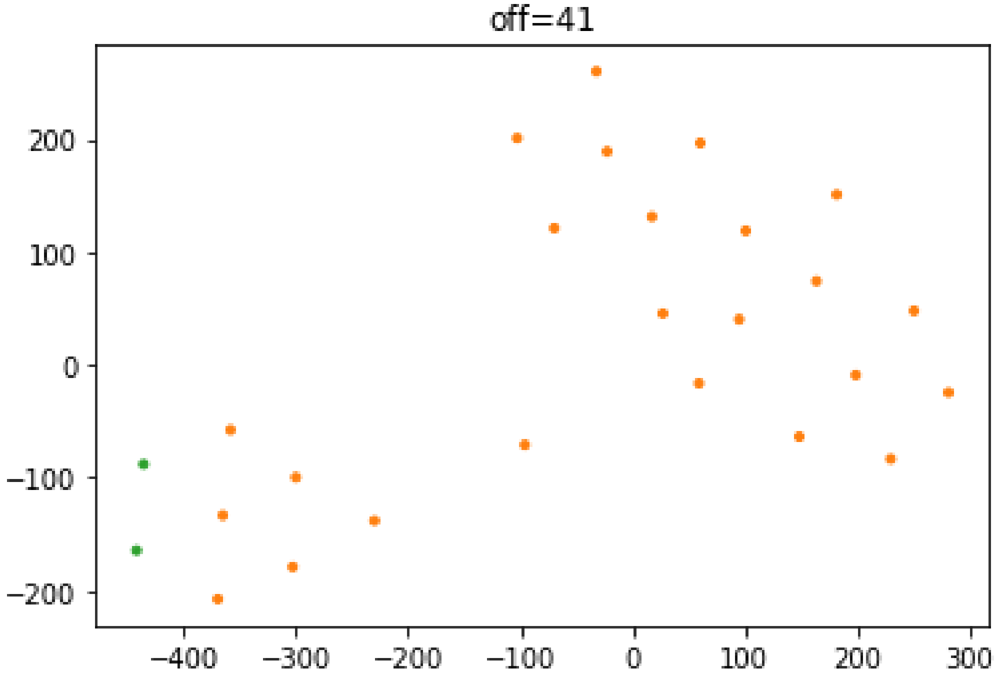

off=36
400
200
0
−200
−400
−600
−400
−200
0
200
400
600
800

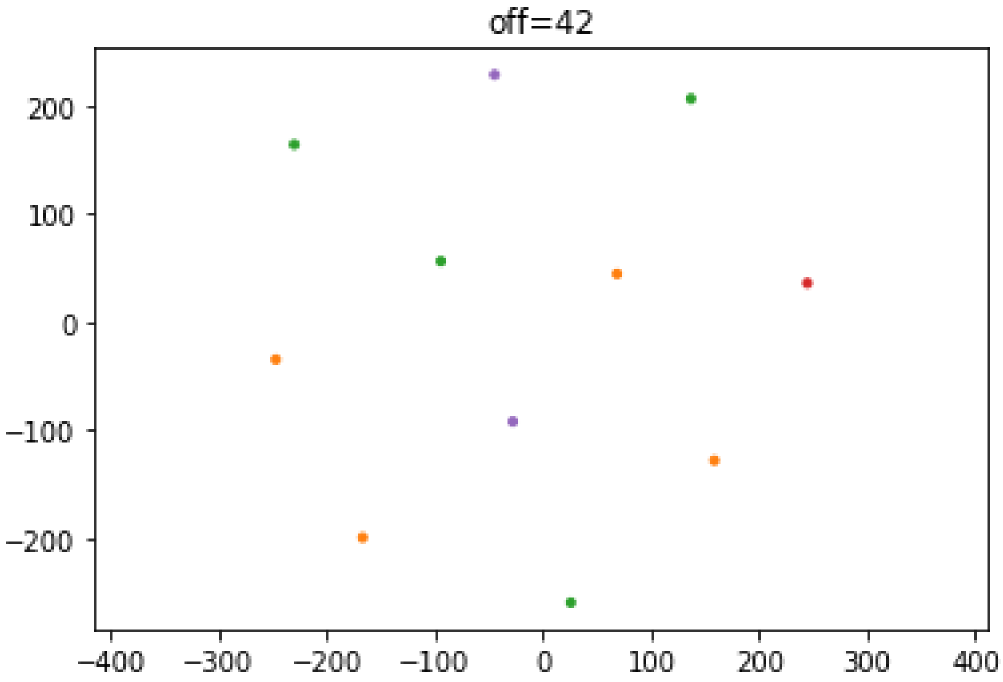

off=37
20
10
0
−10
−20
−30
−40
−20
0
20
40

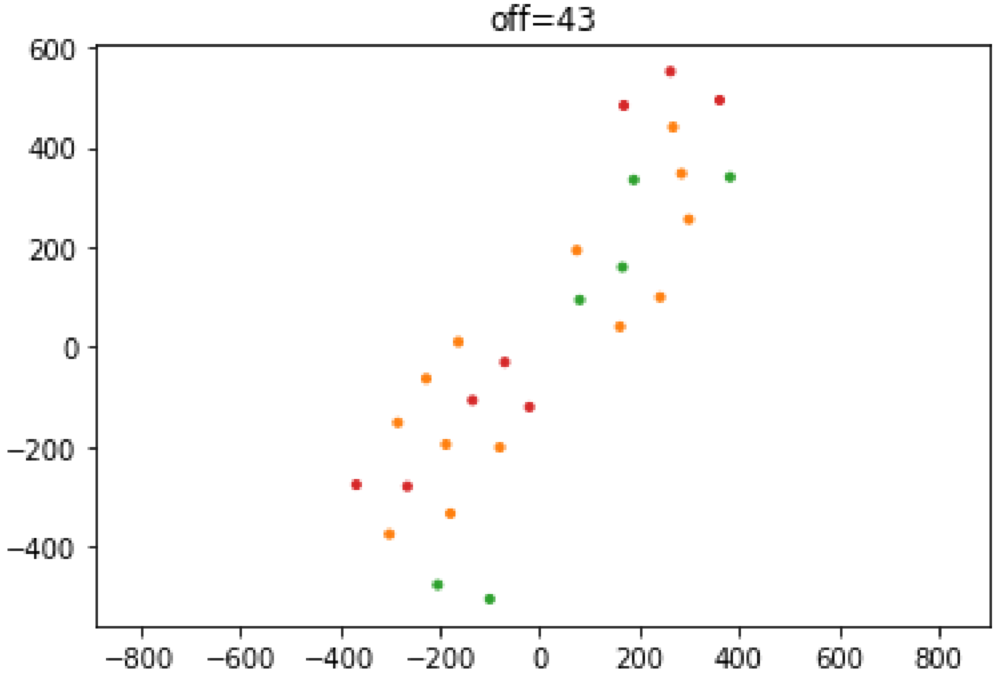

off=38
20
10
0
−10
−20
−30
−20
−10
0
10
20
30
40

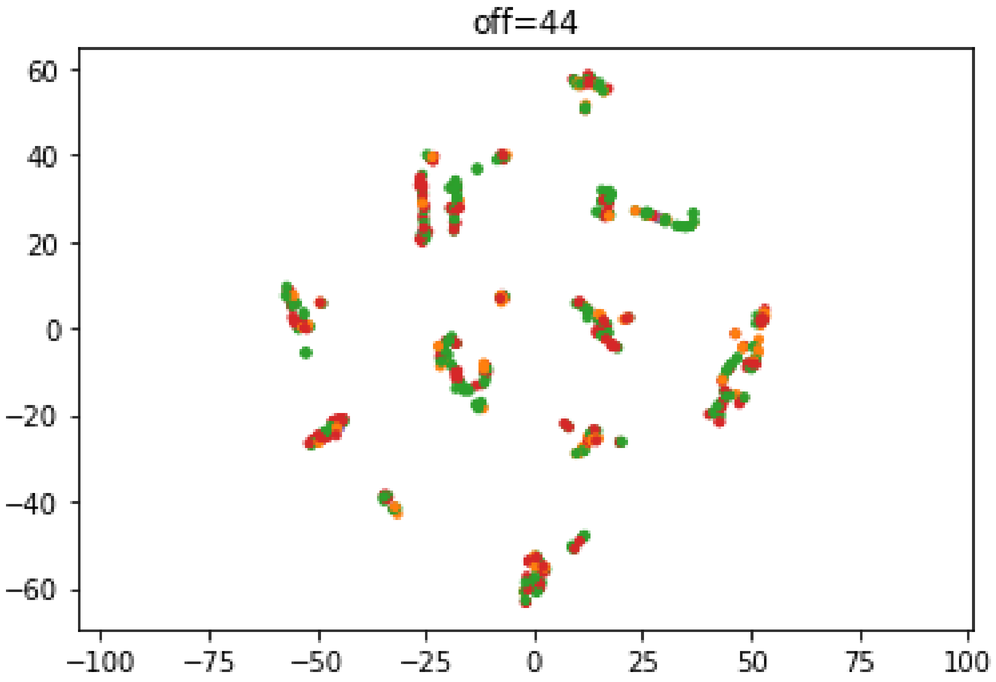

off=39
600
400
200
0
−200
−400
−600
−1000
−750
−500
−250
0
250
500
750
1000

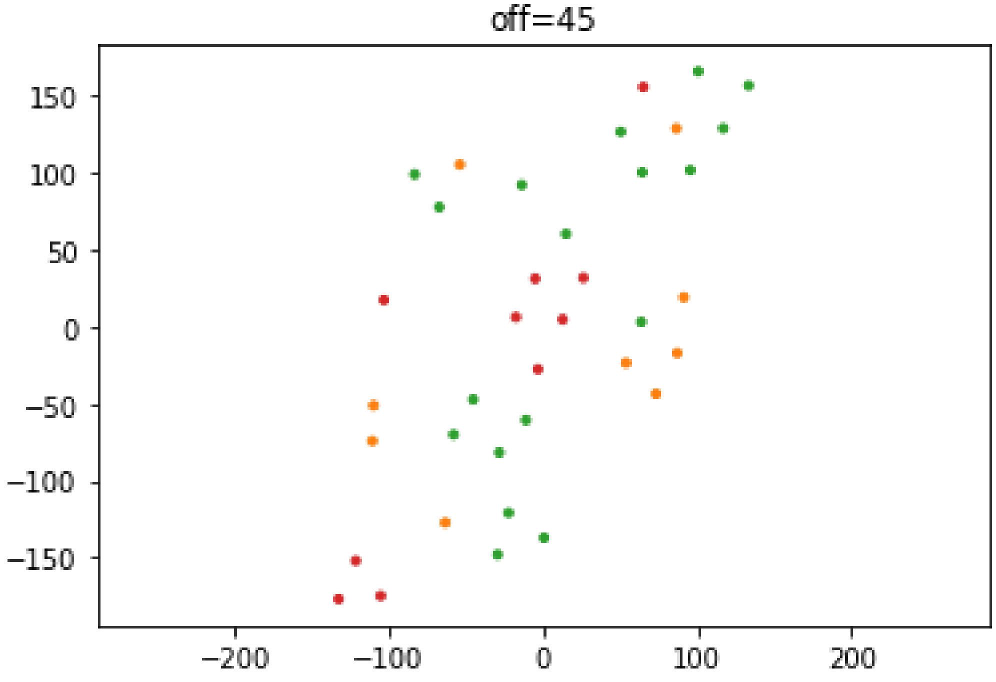

off=40
80
60
40
20
0
−20
−40
−100
−50
0
50

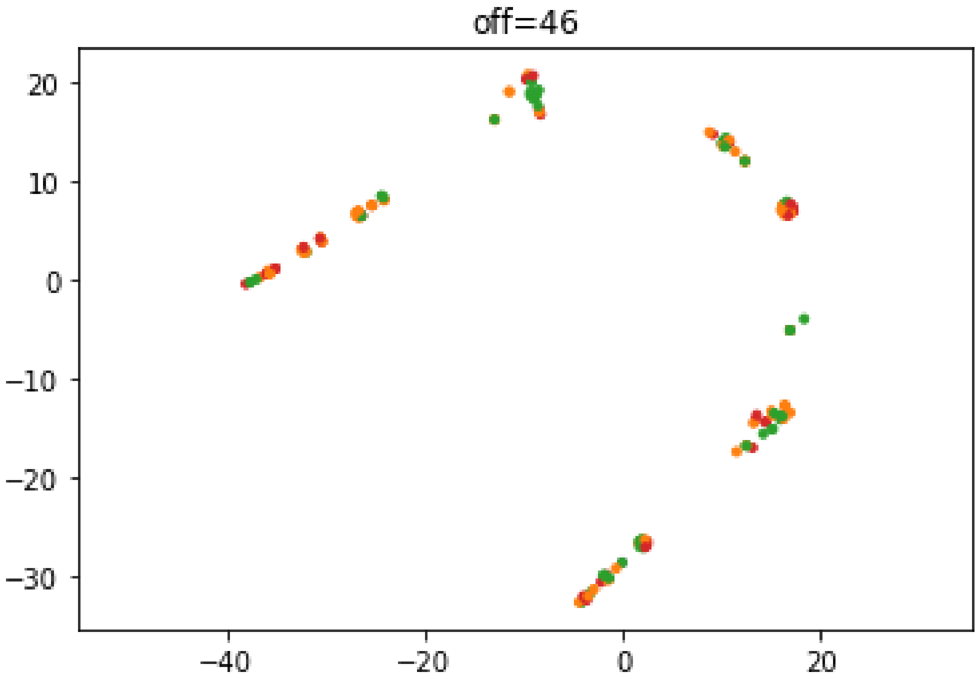

off=41
200
100
0
−100
−200
−400
−300
−200
−100
0
100
200
300

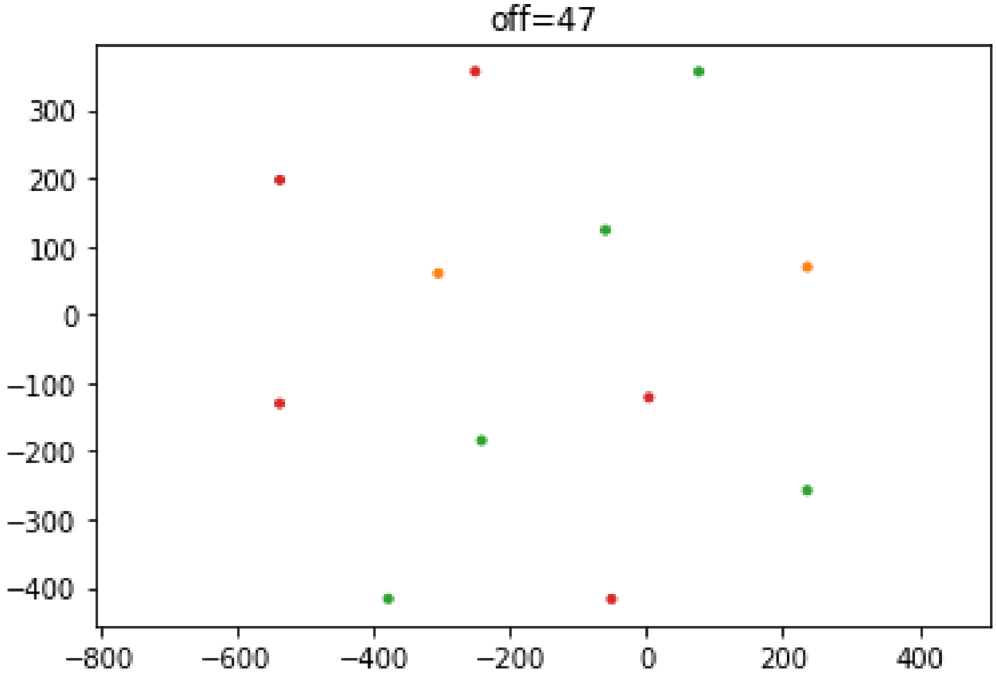

off=42
200
100
0
−100
−200
−400
−300
−200
−100
0
100
200
300
400

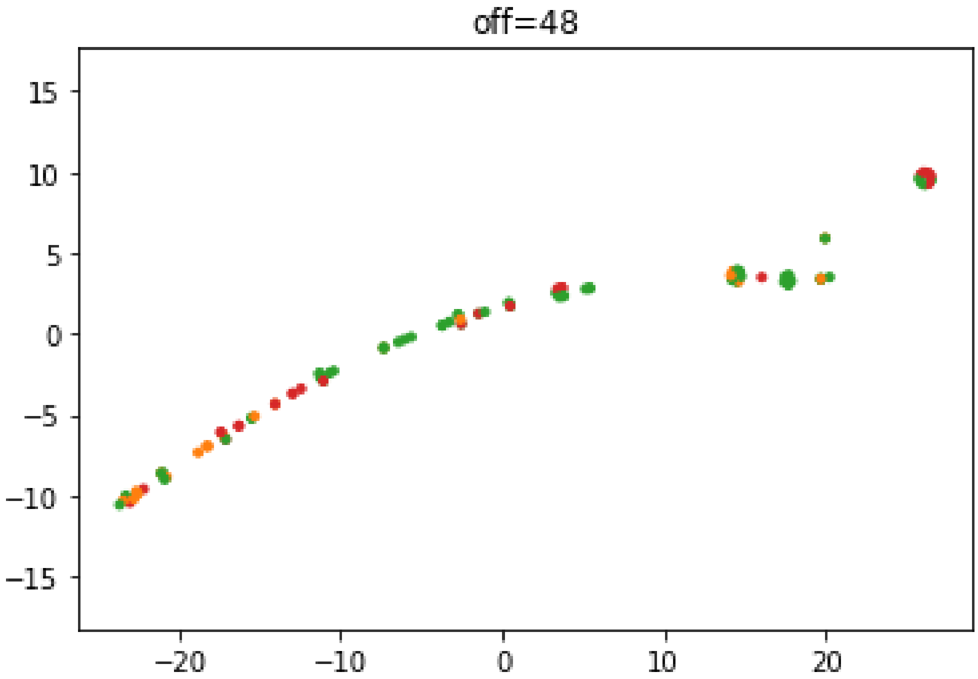

off=43
600
400
200
0
−200
−400
−800
−600
−400
−200
0
200
400
600
800

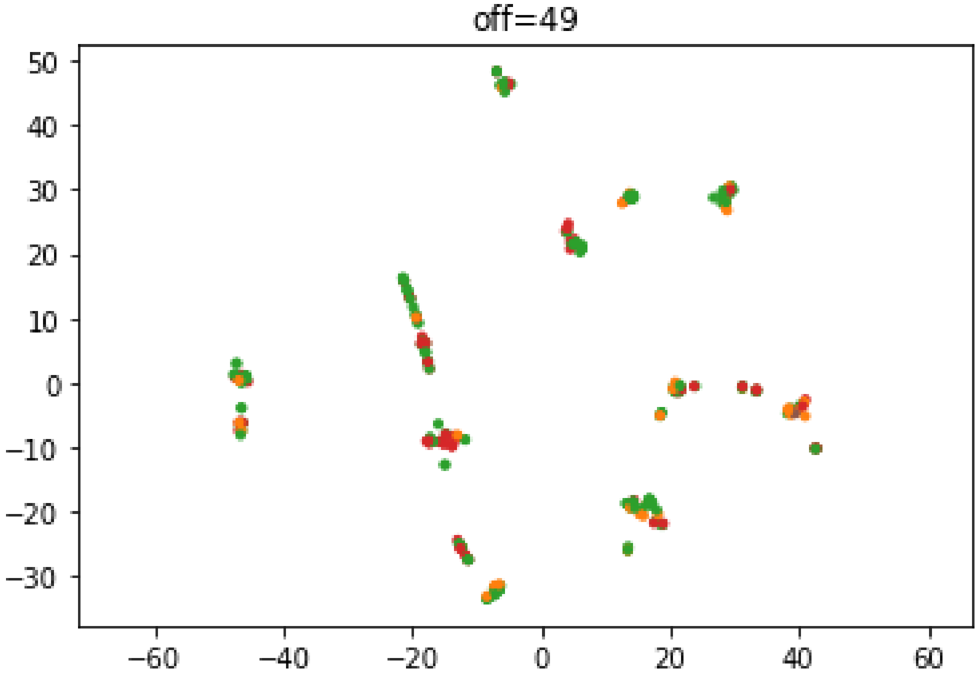

off=44
60
40
20
0
−20
−40
−60
−100
−75
−50
−25
0
25
50
75
100

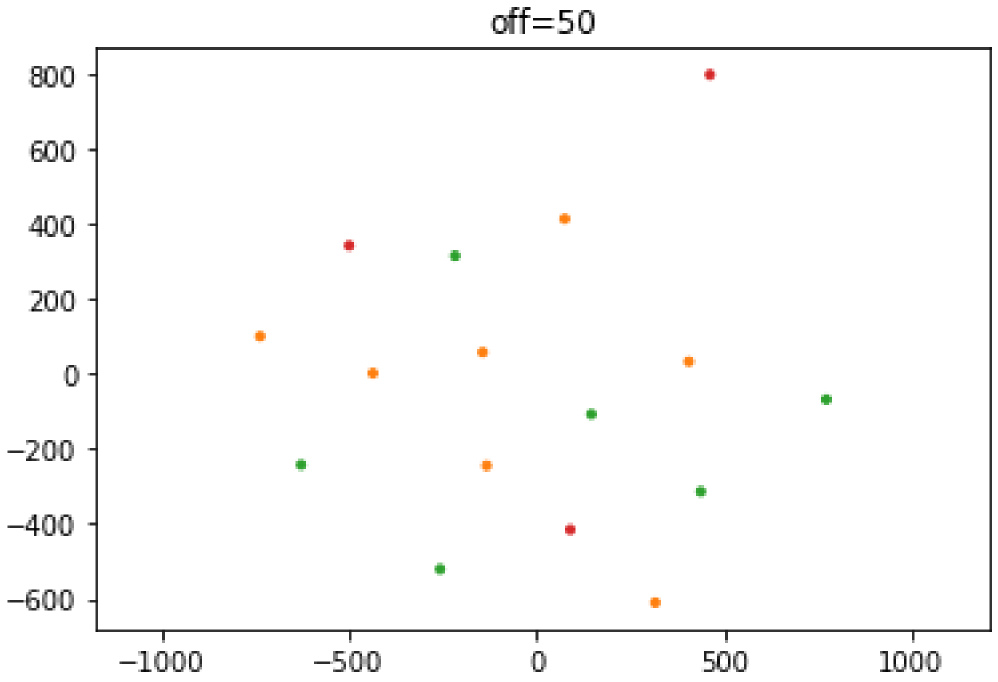

off=45
150
100
50
0
−50
−100
−150
−200
−100
0
100
200

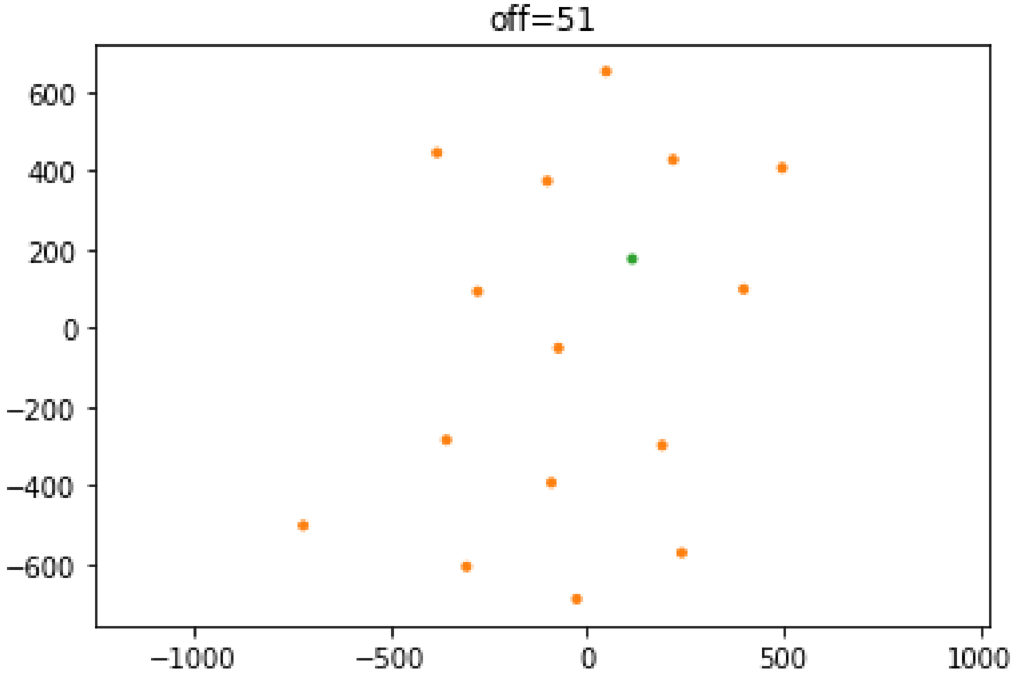

off=46
20
10
0
−10
−20
−30
−40
−20
0
20

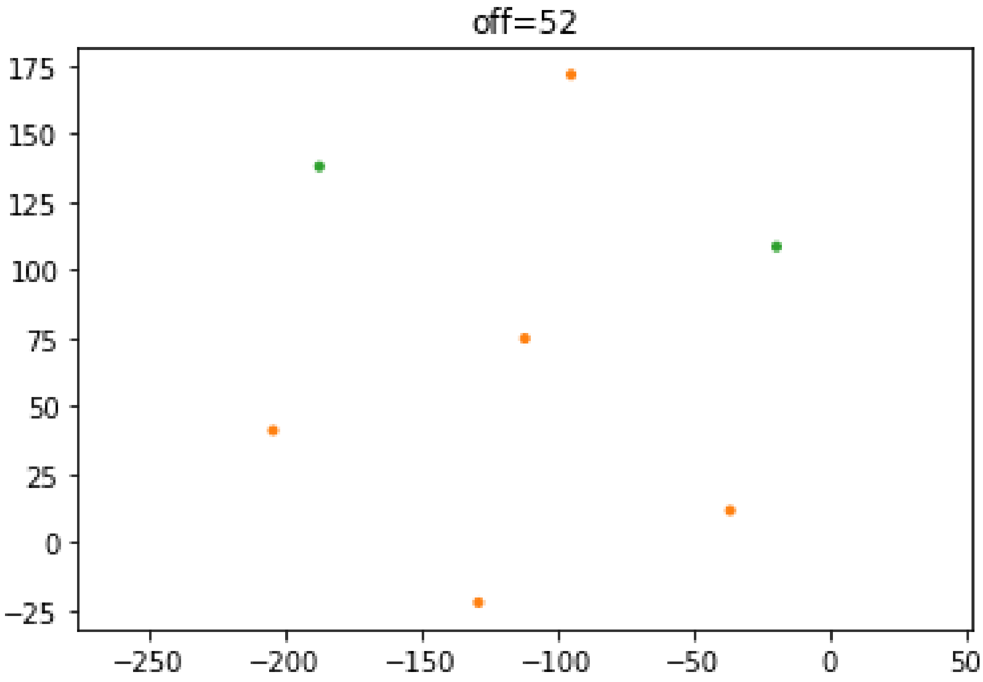

off=47
300
200
100
0
−100
−200
−300
−400
−800
−600
−400
−200
0
200
400

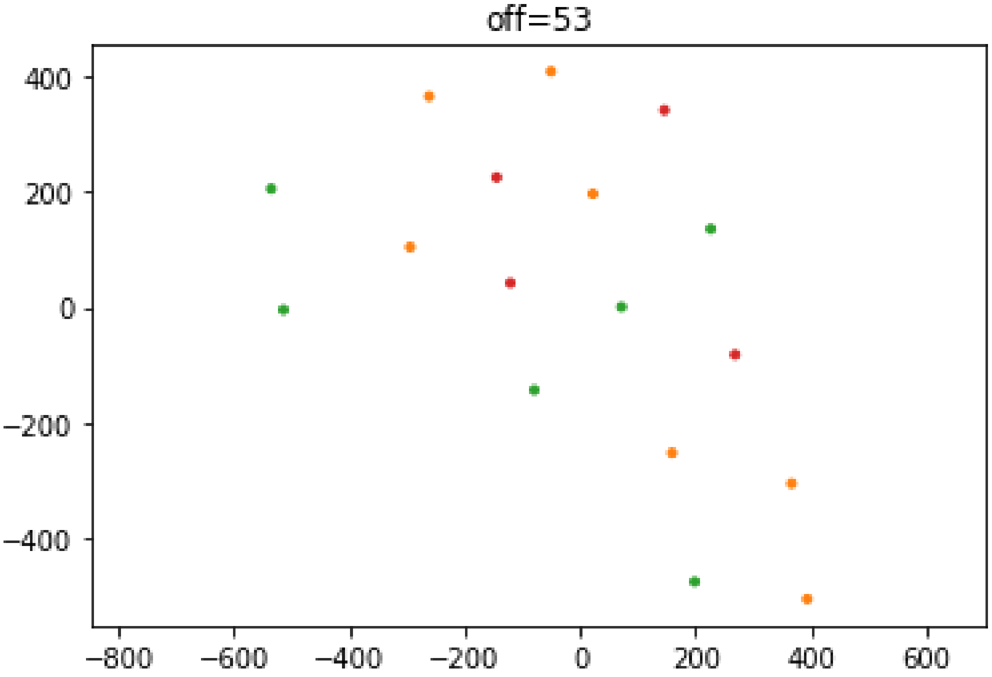

off=48
15
10
5
0
−5
−10
−15
−20
−10
0
10
20

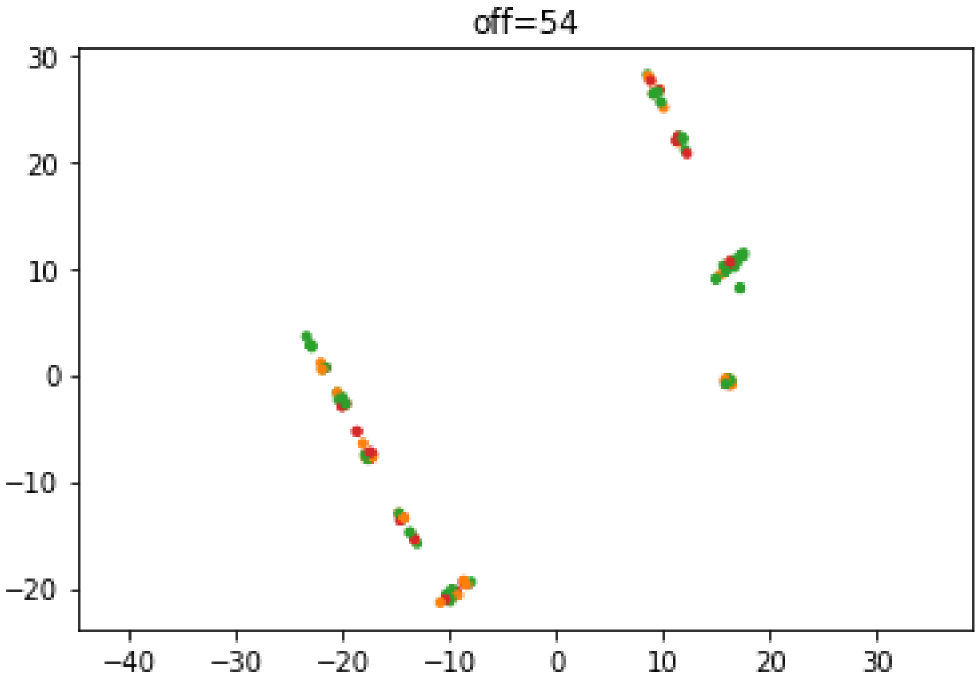

off=49
50
40
30
20
10
0
−10
−20
−30
−60
−40
−20
0
20
40
60

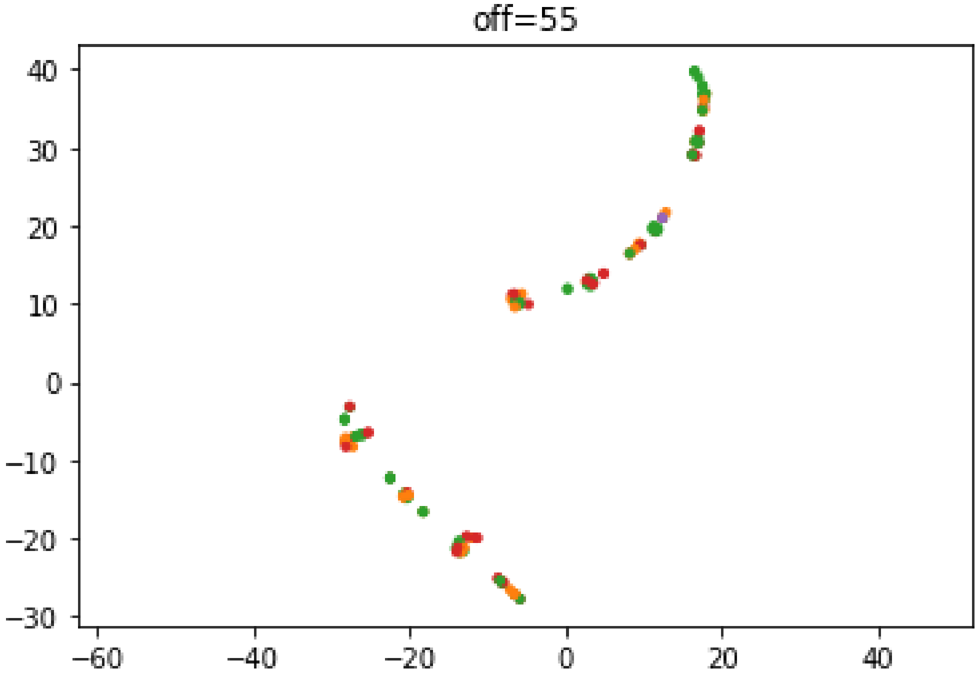

off=50
800
600
400
200
0
−200
−400
−600
−1000
−500
0
500
1000

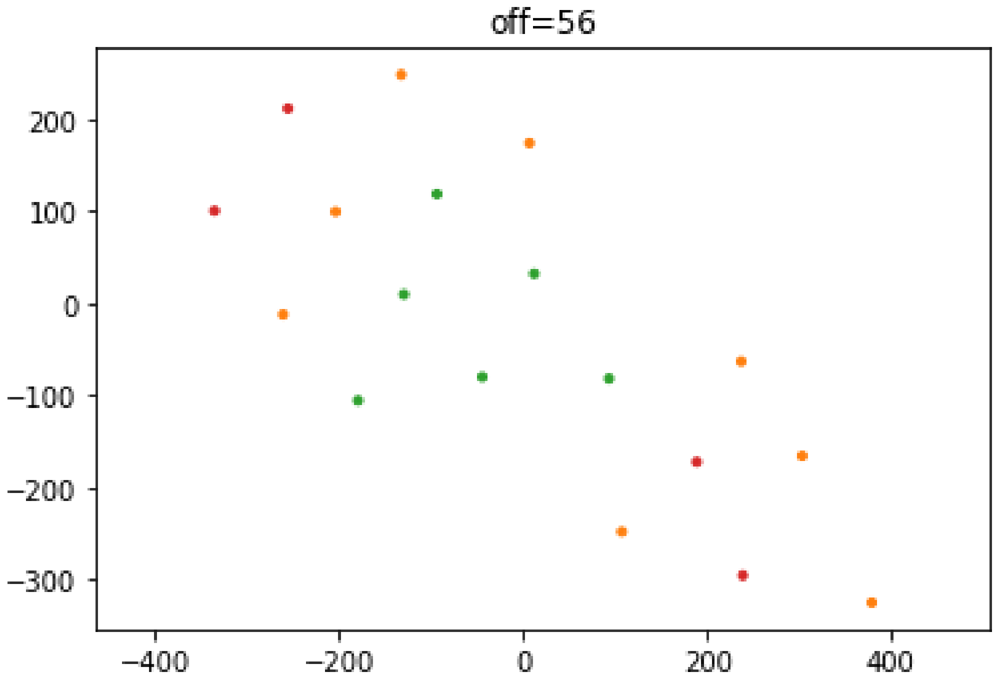

off=51
600
400
200
0
−200
−400
−600
−1000
−500
0
500
1000

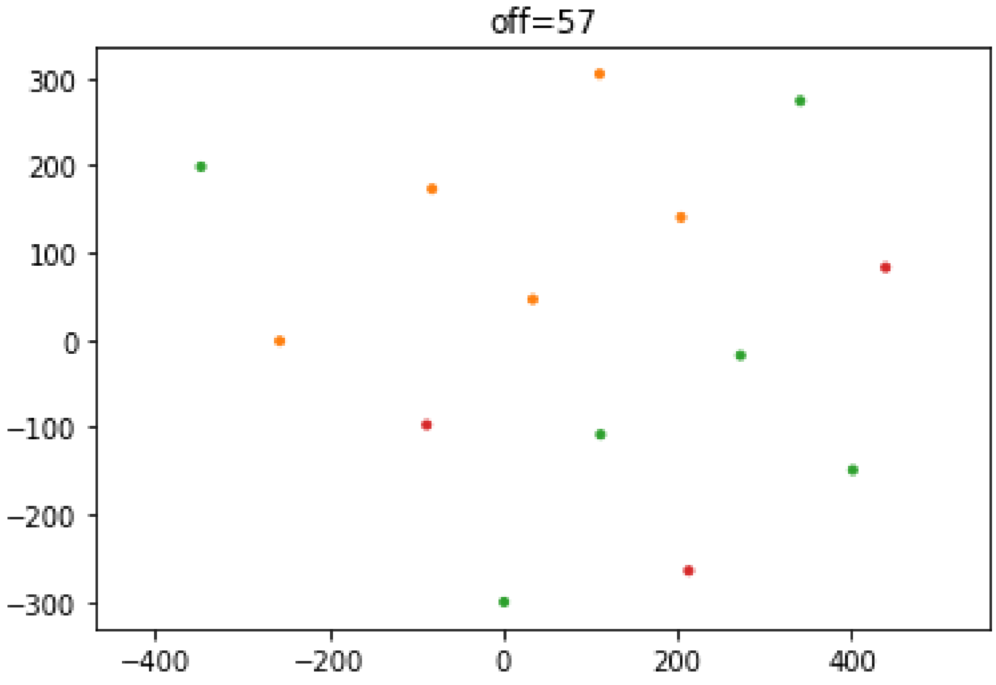

off=52
175
150
125
100
75
50
25
0
−25
−250
−200
−150
−100
−50
0
50

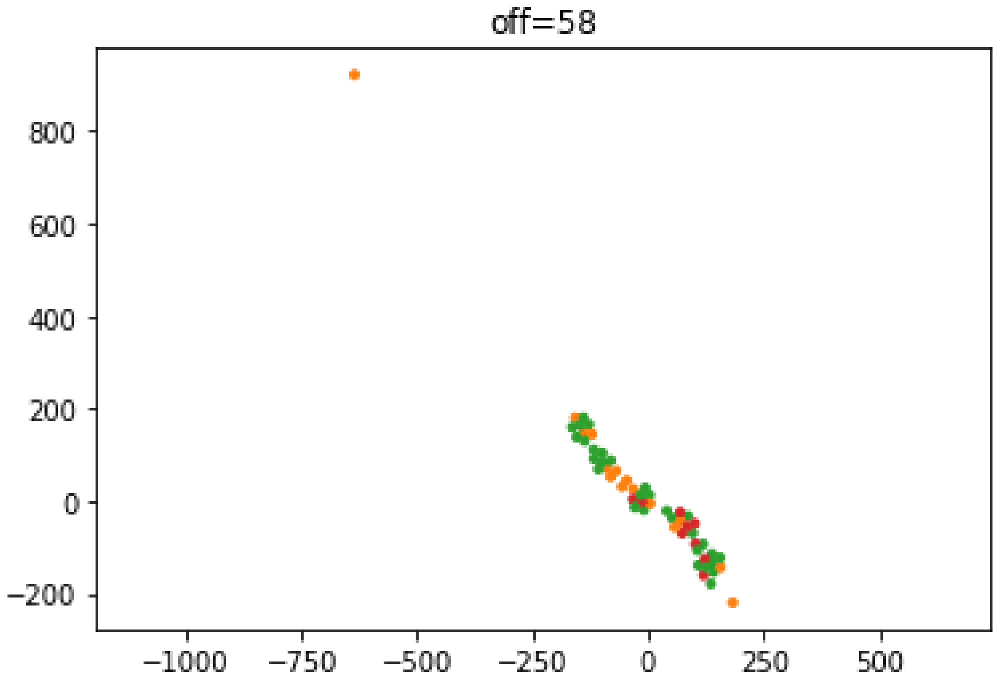

off=53
400
200
0
−200
−400
−800
−600
−400
−200
0
200
400
600

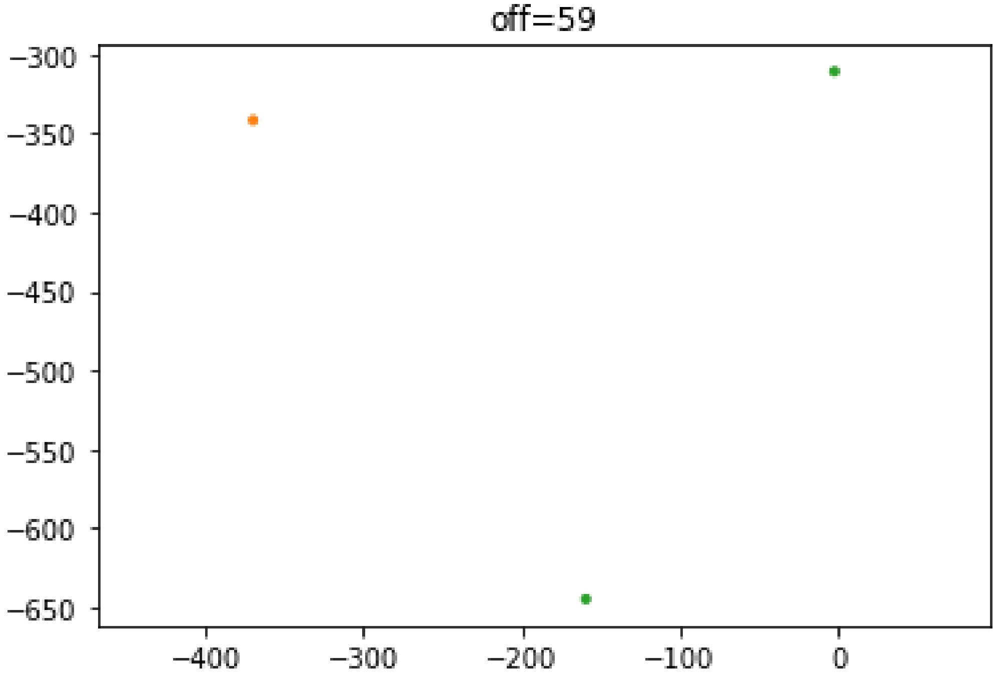

off=54
30
20
10
0
−10
−20
−40
−30
−20
−10
0
10
20
30

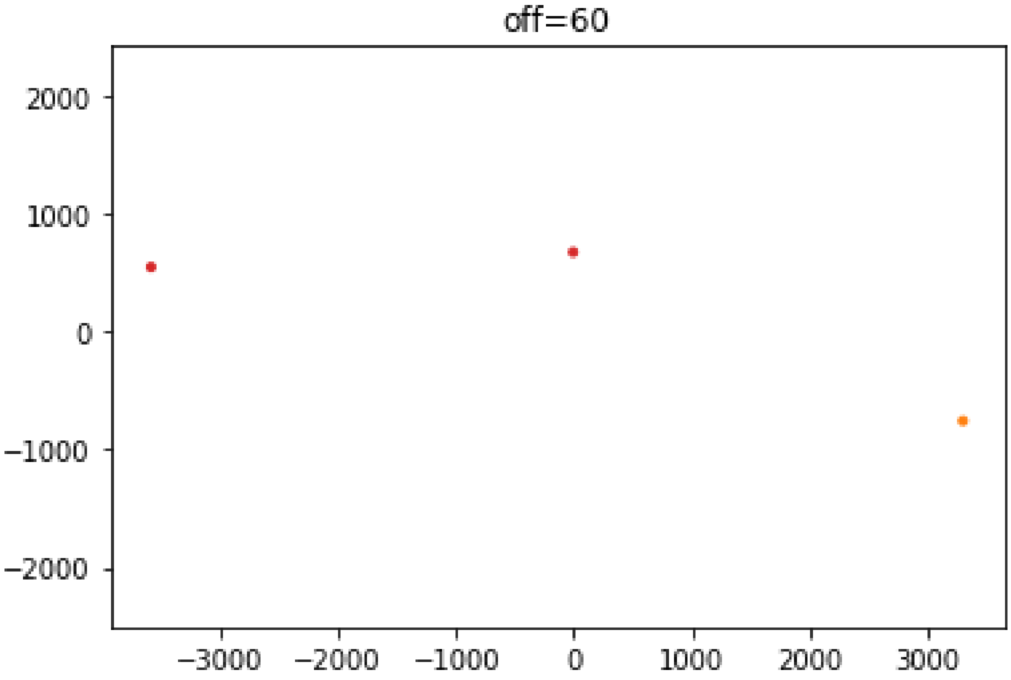

off=55
40
30
20
10
0
−10
−20
−30
−60
−40
−20
0
20
40

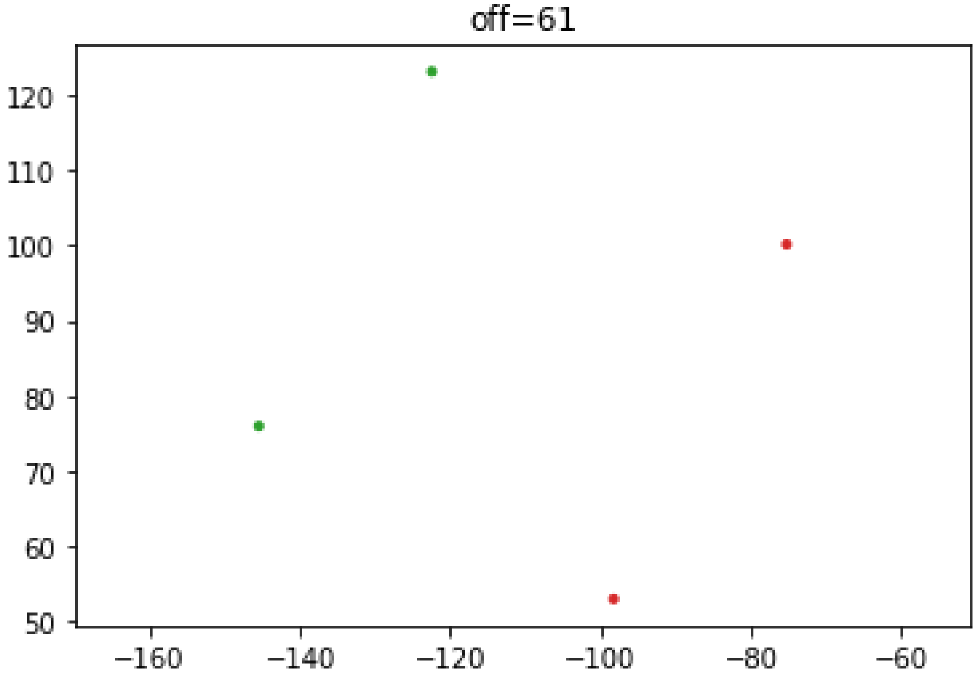

off=56
200
100
0
−100
−200
−300
−400
−200
0
200
400

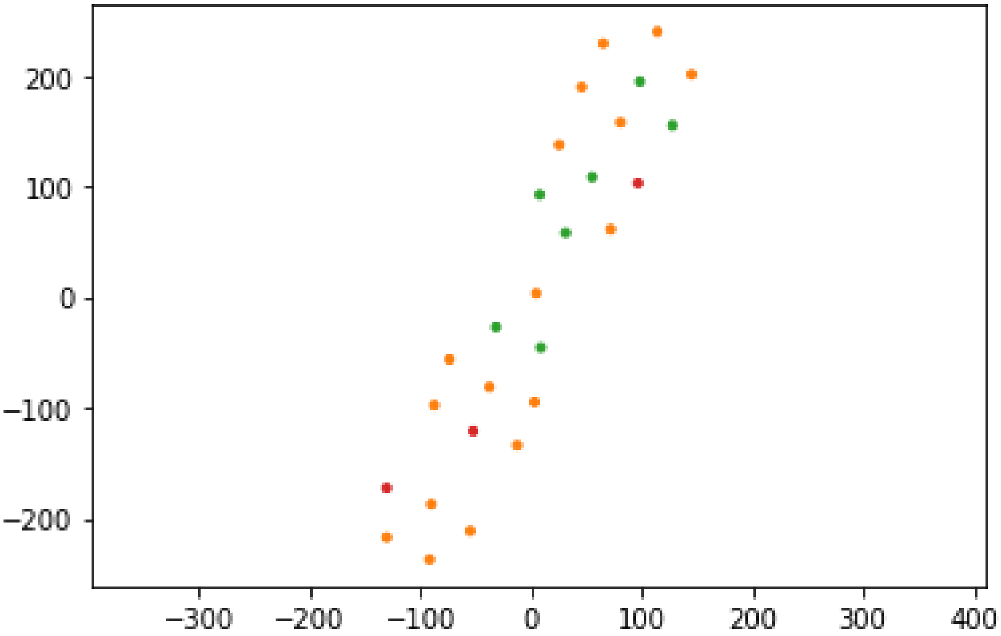

off=57
300
200
100
0
−100
−200
−300
−400
−200
0
200
400


off=58
800
600
400
200
0
−200
−1000
−750
−500
−250
0
250
500

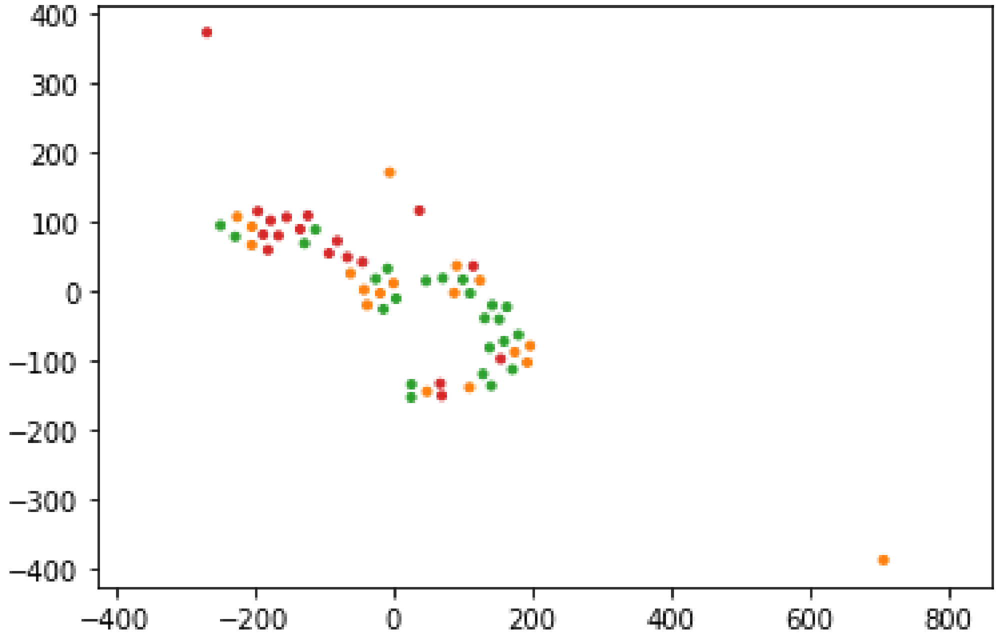

off=59
−300
−350
−400
−450
−500
−550
−600
−650
−400
−300
−200
−100
0


off=60
2000
1000
0
−1000
−2000
−3000
−2000
−1000
0
1000
2000
3000

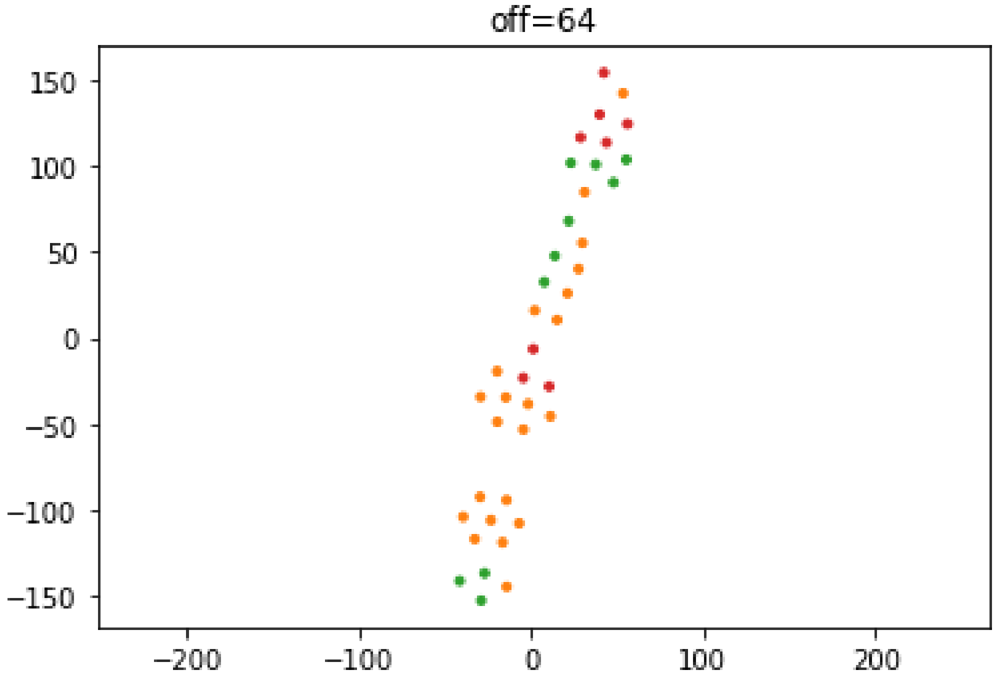

off=61
120
110
100
90
80
70
60
50
−160
−140
−120
−100
−80
−60

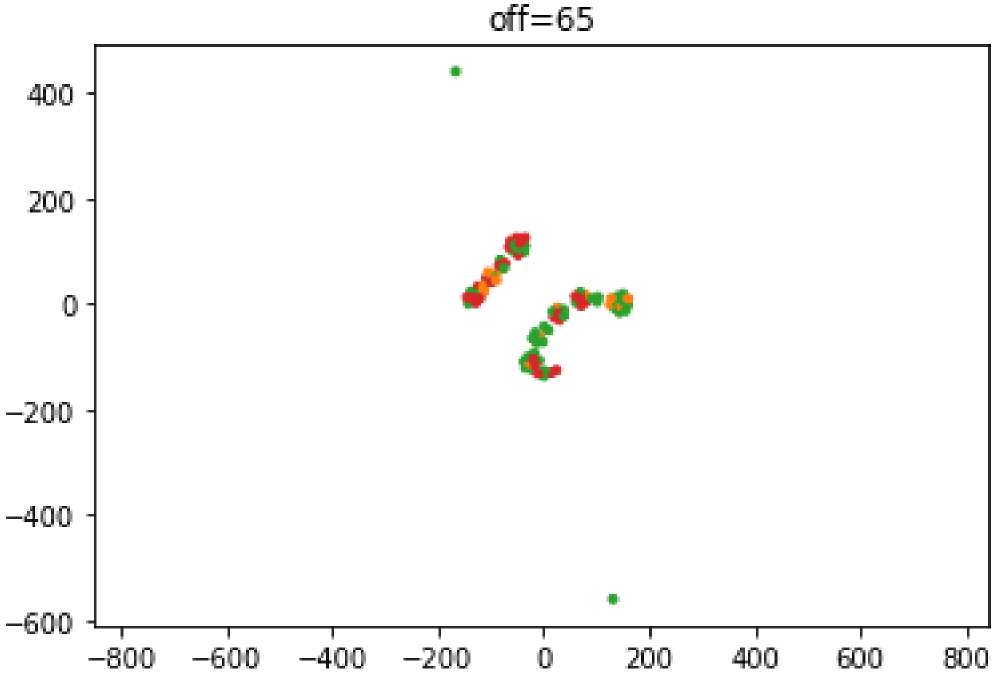

off=62

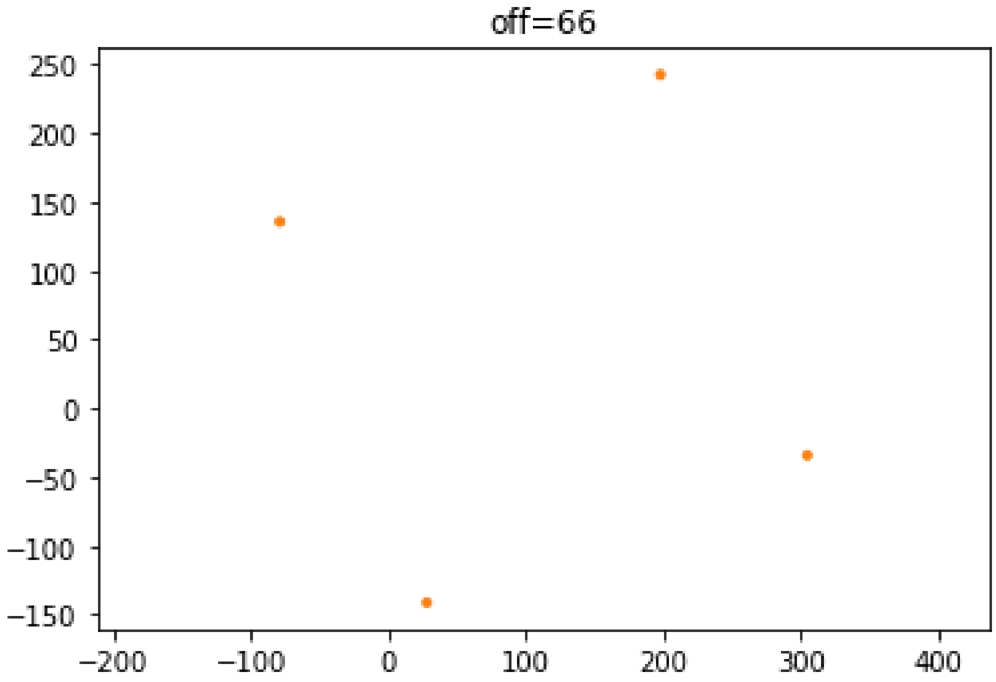
200
100
0
−100
−200
−300
−200
−100
0
100
200
300
400

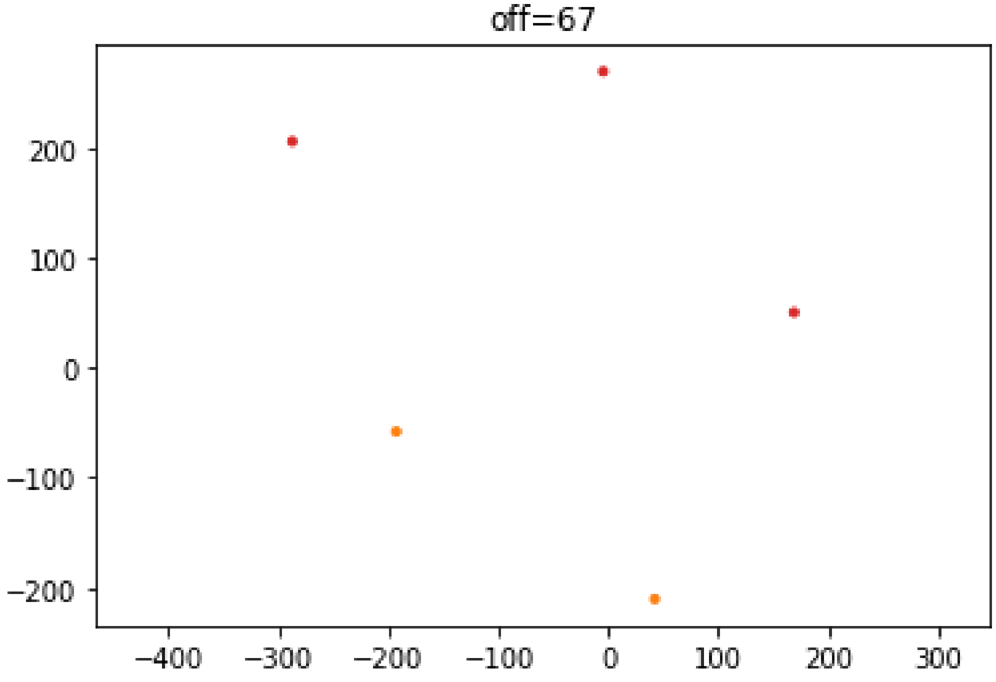

off=63

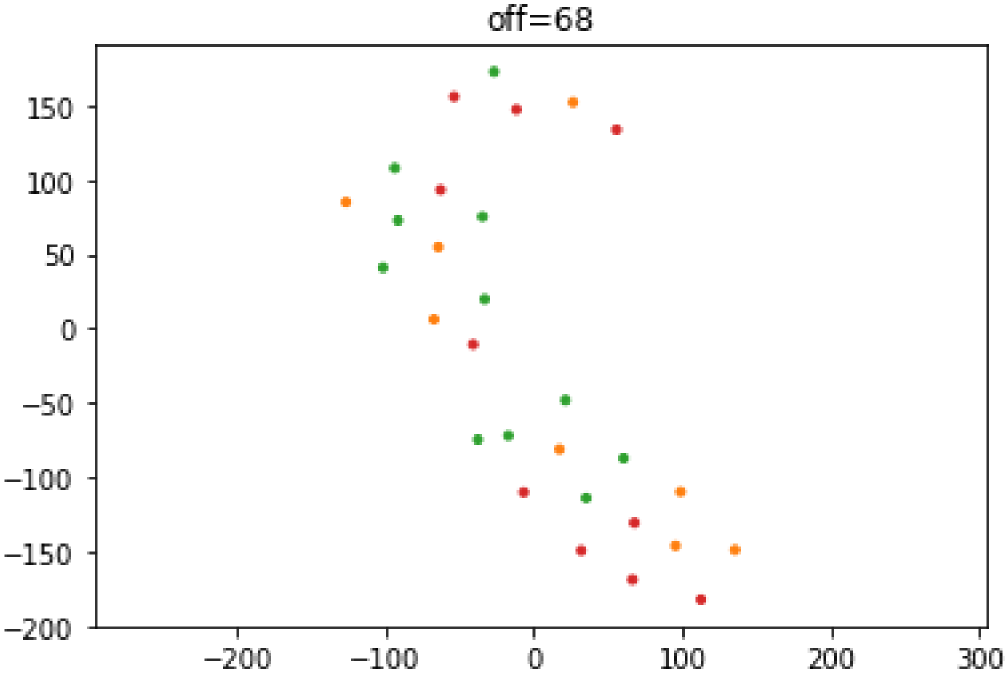
400
300
200
100
0
−100
−200
−300
−400
−400
−200
0
200
400
600
800

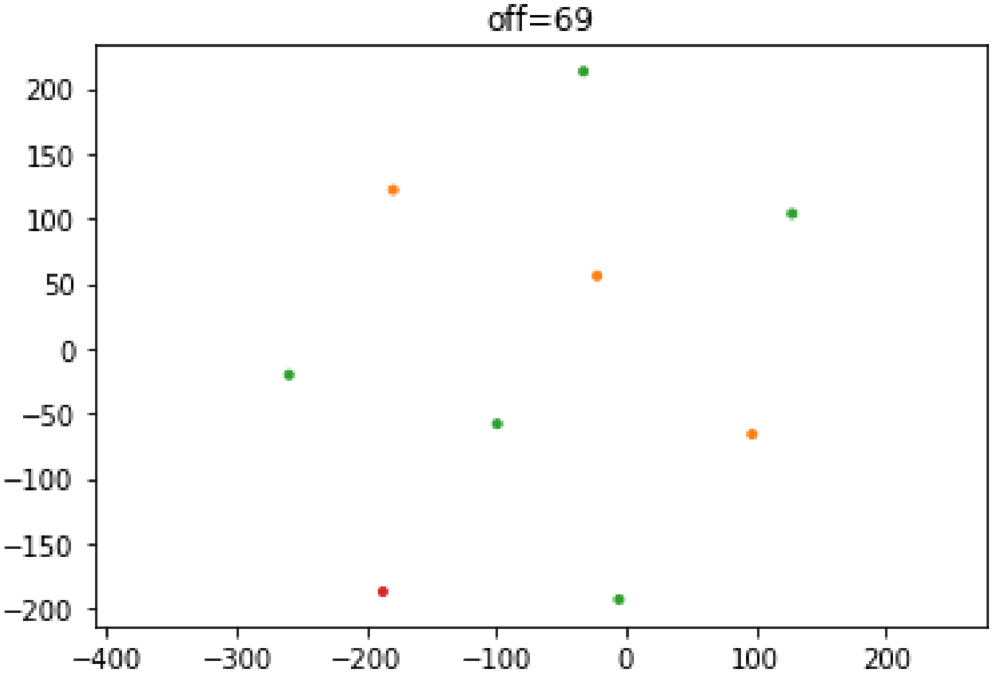

off=64
150
100
50
0
−50
−100
−150
−200
−100
0
100
200

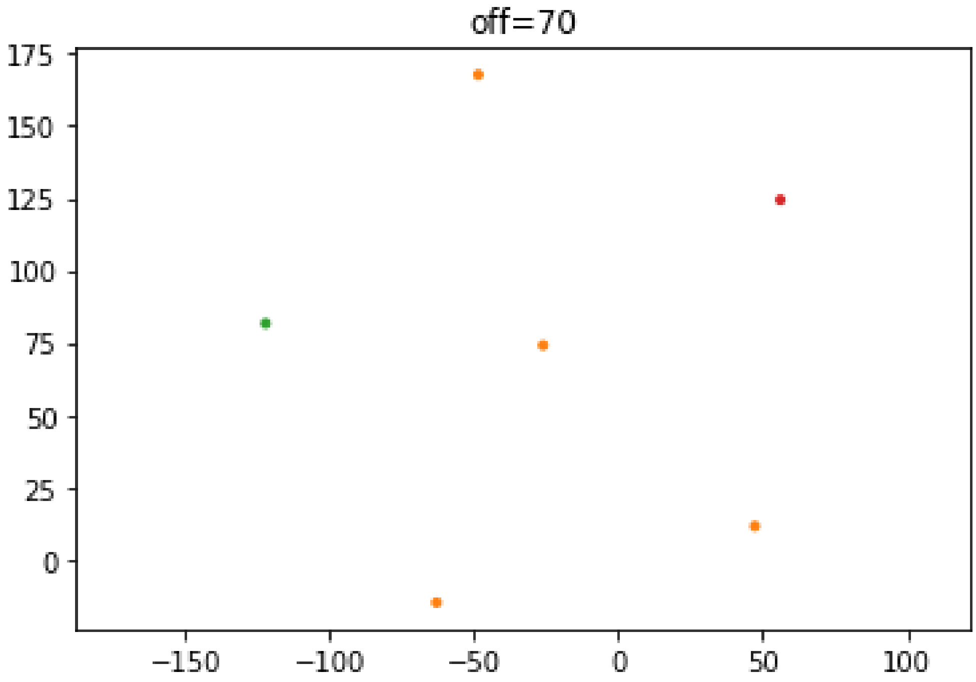

off=65
400
200
0
−200
−400
−600
−800
−600
−400
−200
0
200
400
600
800

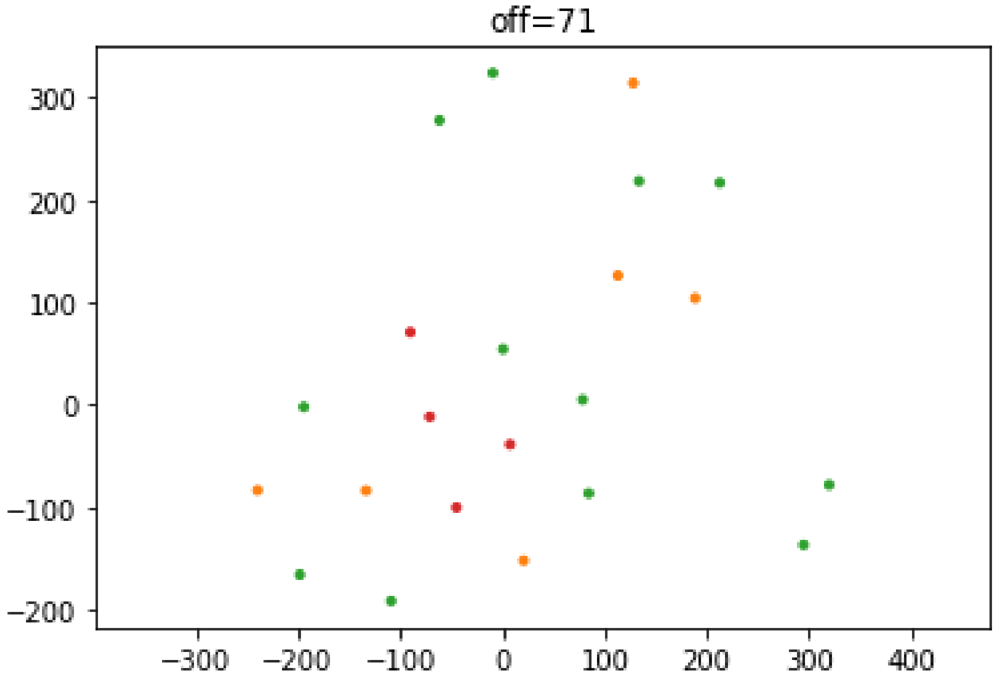

off=66
250
200
150
100
50
0
−50
−100
−150
−200
−100
0
100
200
300
400

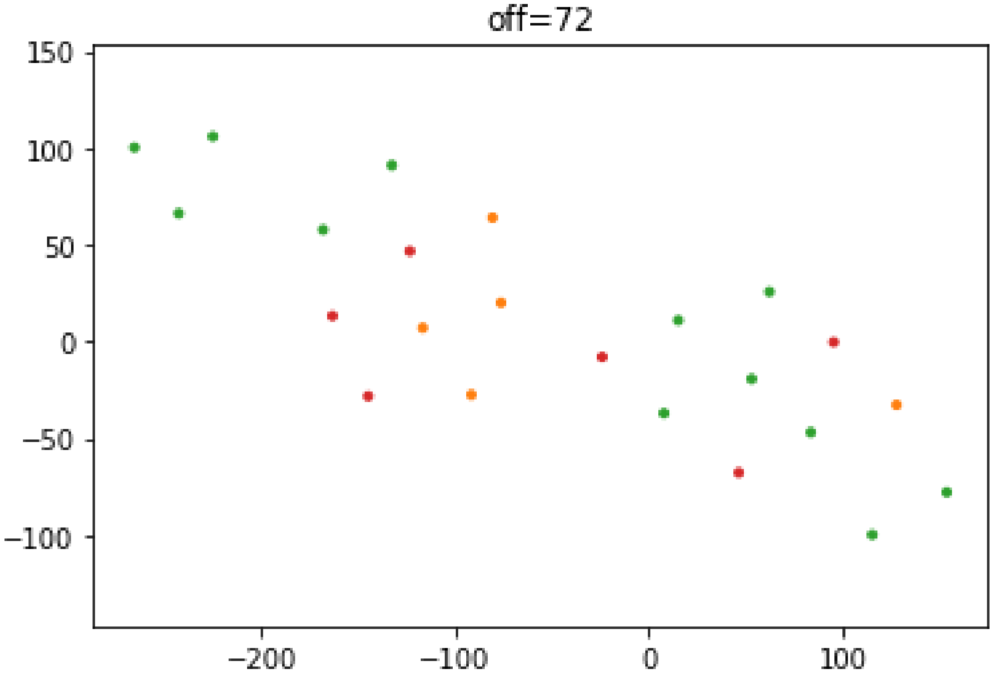

off=67
200
100
0
−100
−200
−400
−300
−200
−100
0
100
200
300

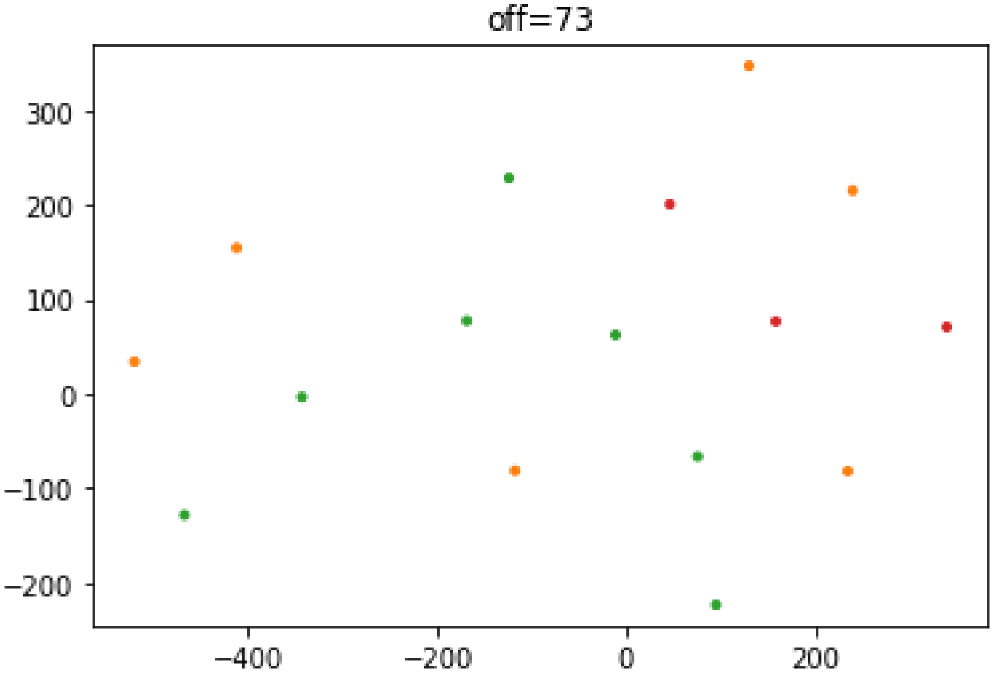

off=68
150
100
50
0
−50
−100
−150
−200
−200
−100
0
100
200
300

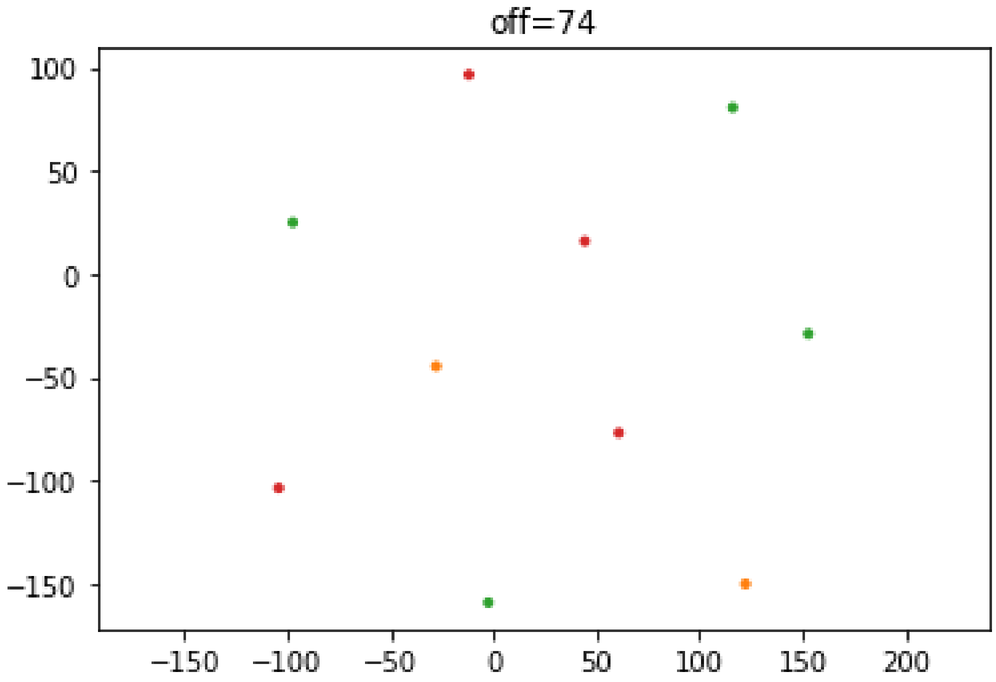

off=69
200
150
100
50
0
−50
−100
−150
−200
−400
−300
−200
−100
0
100
200

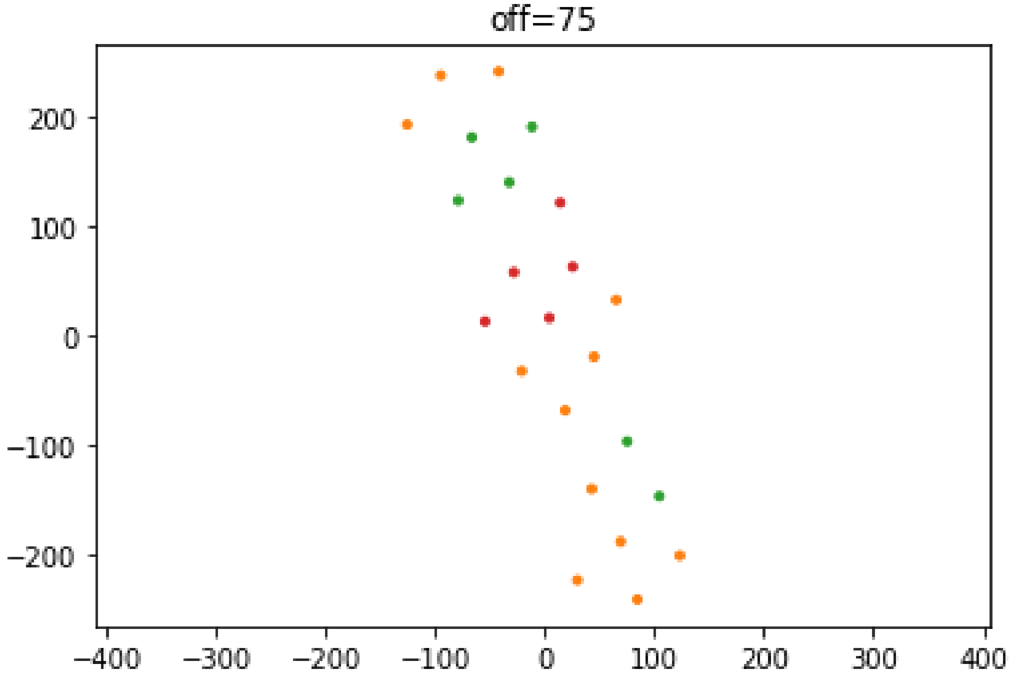

off=70
175
150
125
100
75
50
25
0
−150
−100
−50
0
50
100

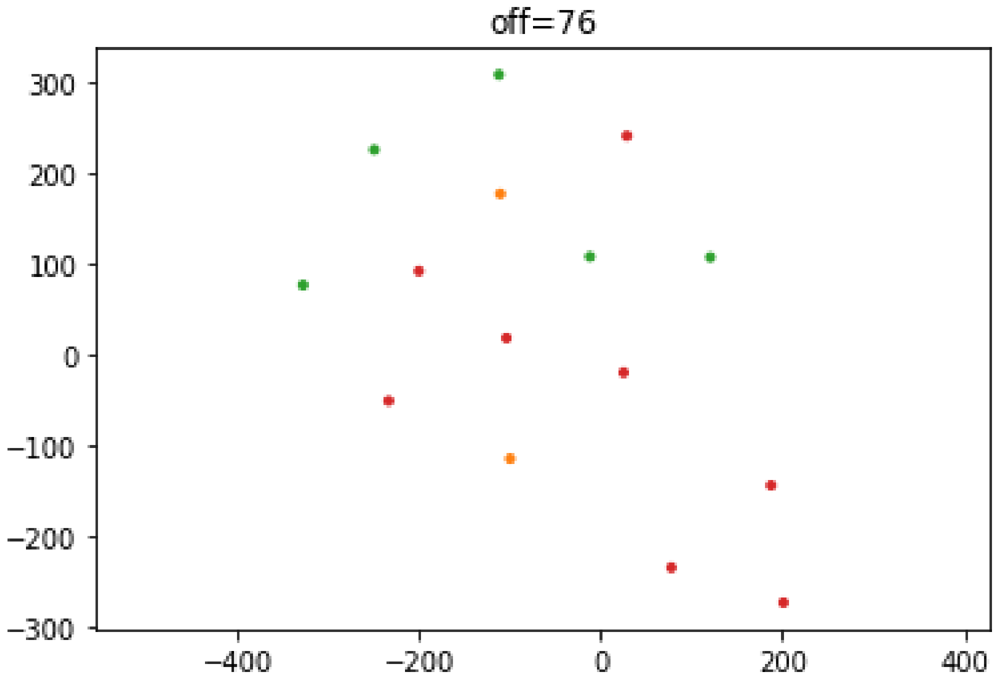

off=71
300
200
100
0
−100
−200
−300
−200
−100
0
100
200
300
400

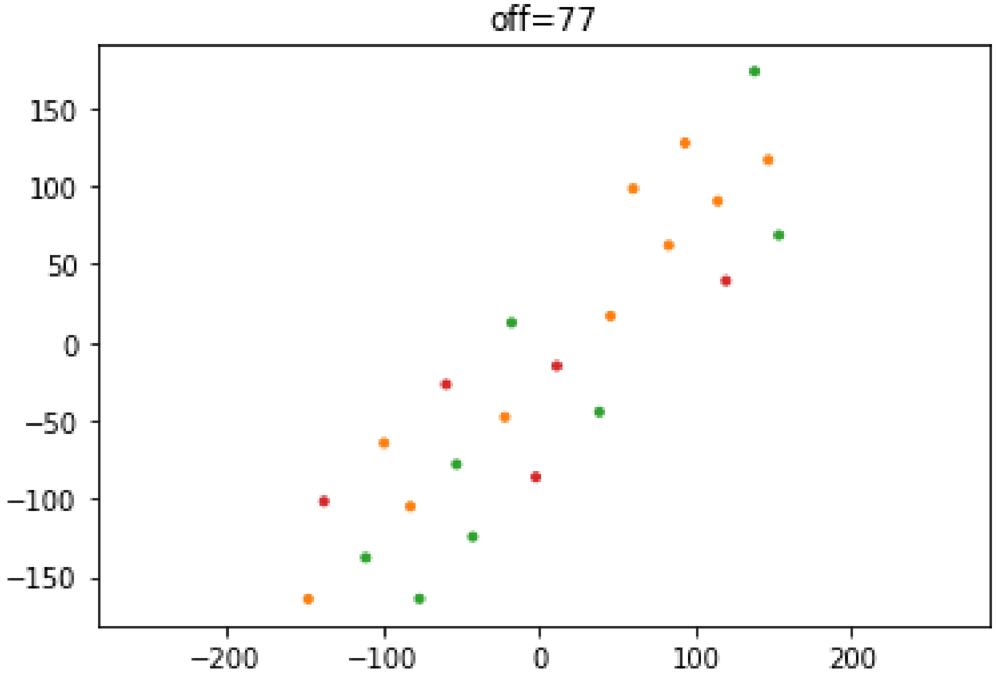

off=72
150
100
50
0
−50
−100
−200
−100
0
100

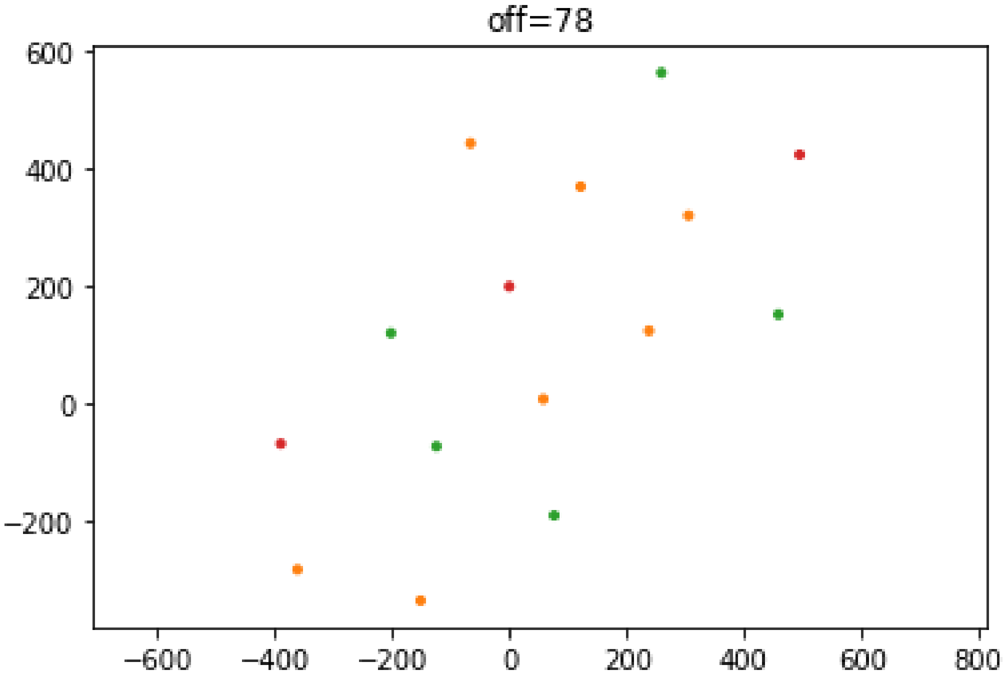

off=73
300
200
100
0
−100
−200
−400
−200
0
200

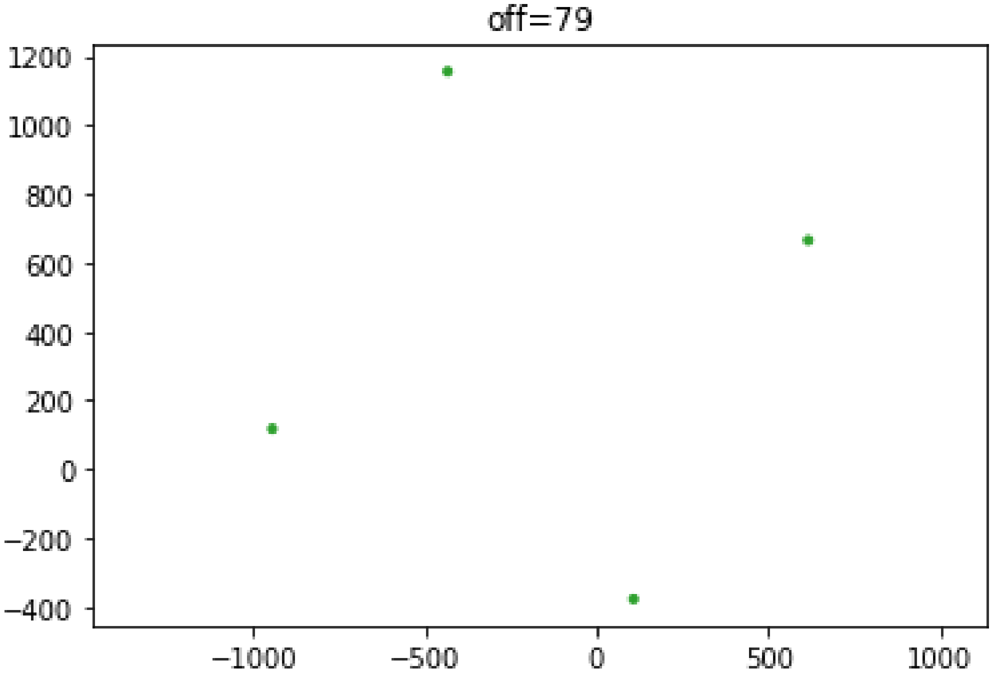

off=74
100
50
0
−50
−100
−150
−150
−100
−50
0
50
100
150
200

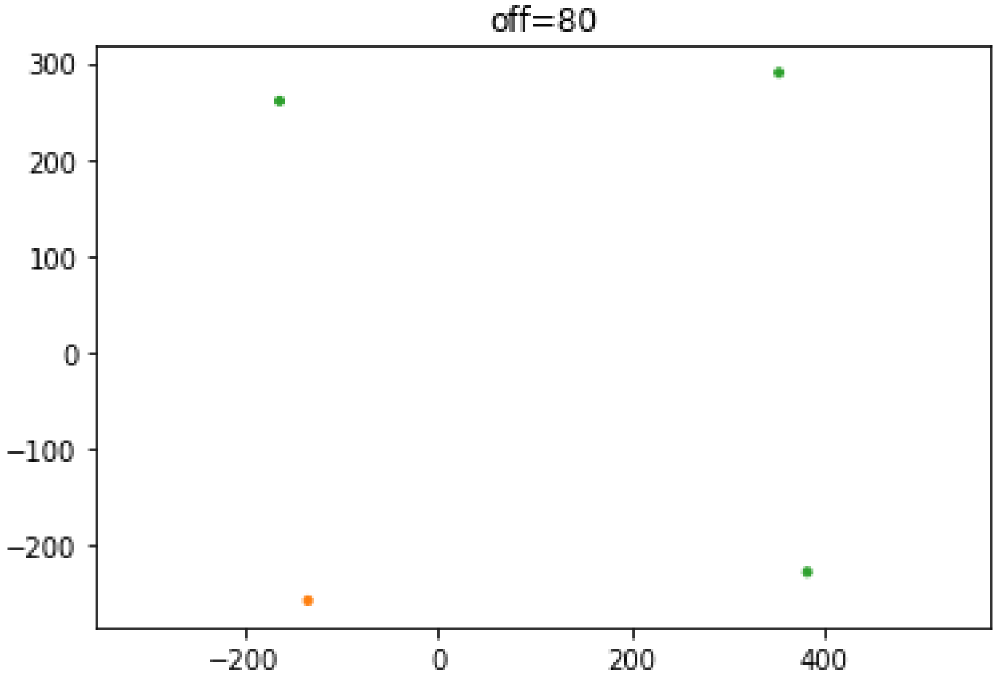

off=75
200
100
0
−100
−200
−400
−300
−200
−100
0
100
200
300
400

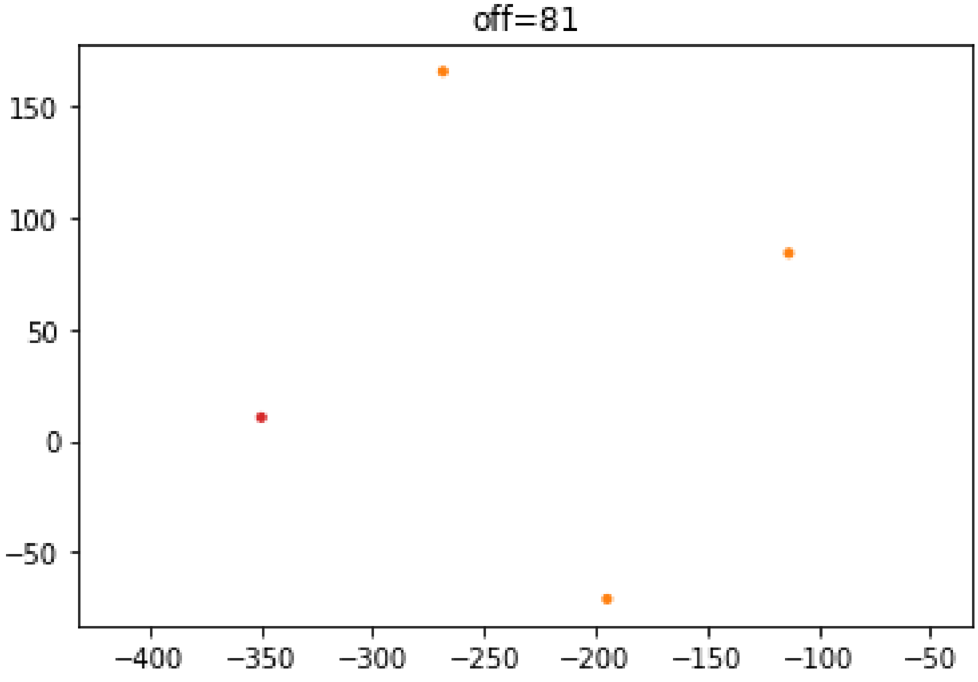

off=76
300
200
100
0
−100
−200
−300
−400
−200
0
200
400

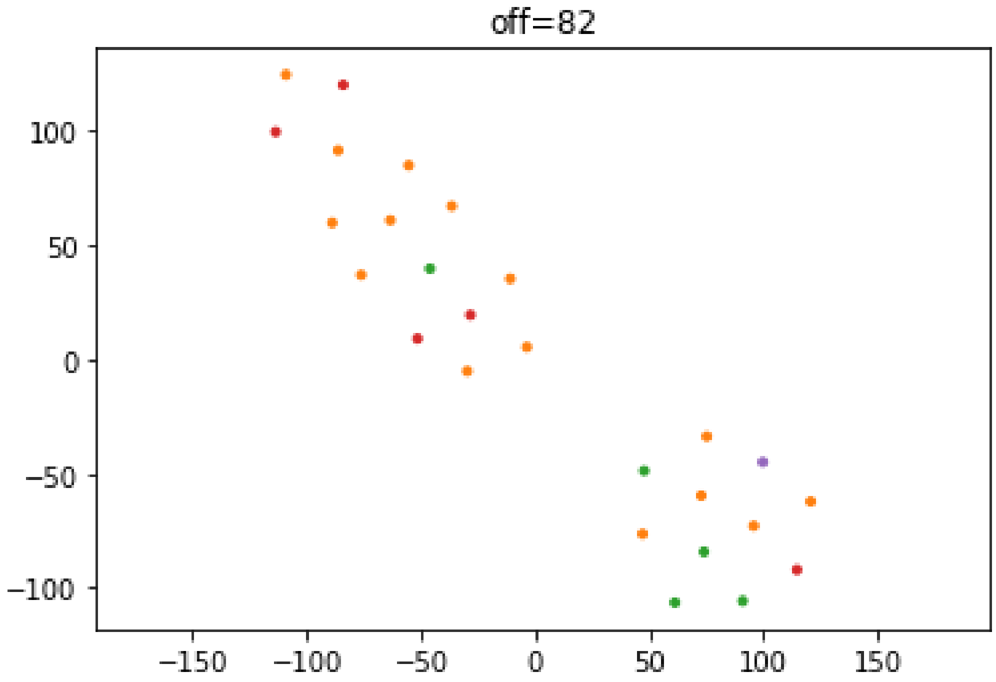

off=77
150
100
50
0
−50
−100
−150
−200
−100
0
100
200

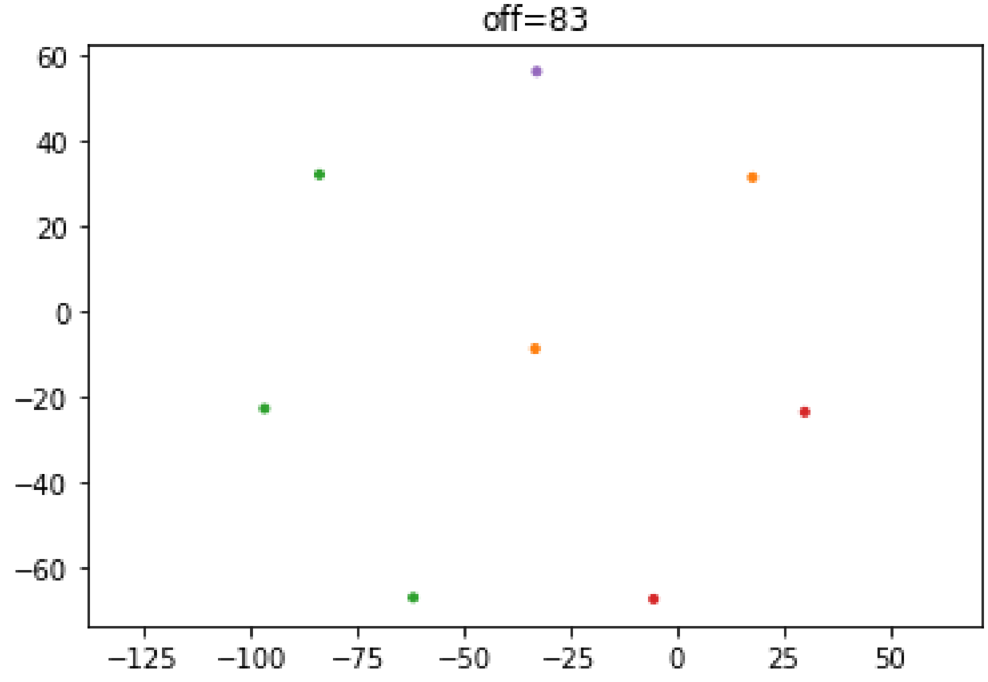

off=78
600
400
200
0
−200
−600
−400
−200
0
200
400
600
800

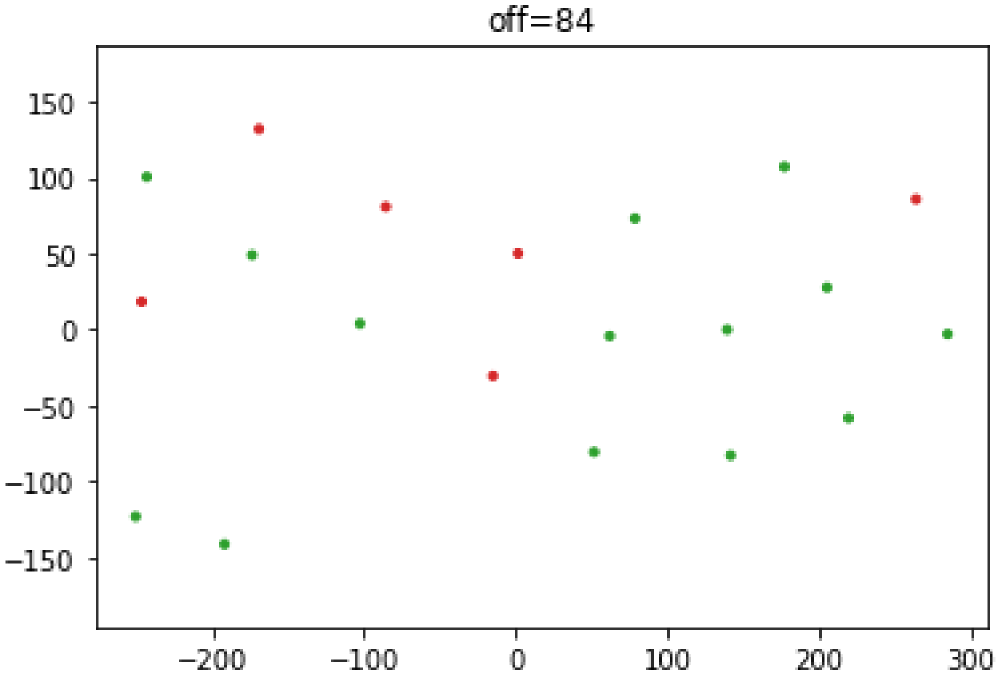

off=79
1200
1000
800
600
400
200
0
−200
−400
−1000
−500
0
500
1000

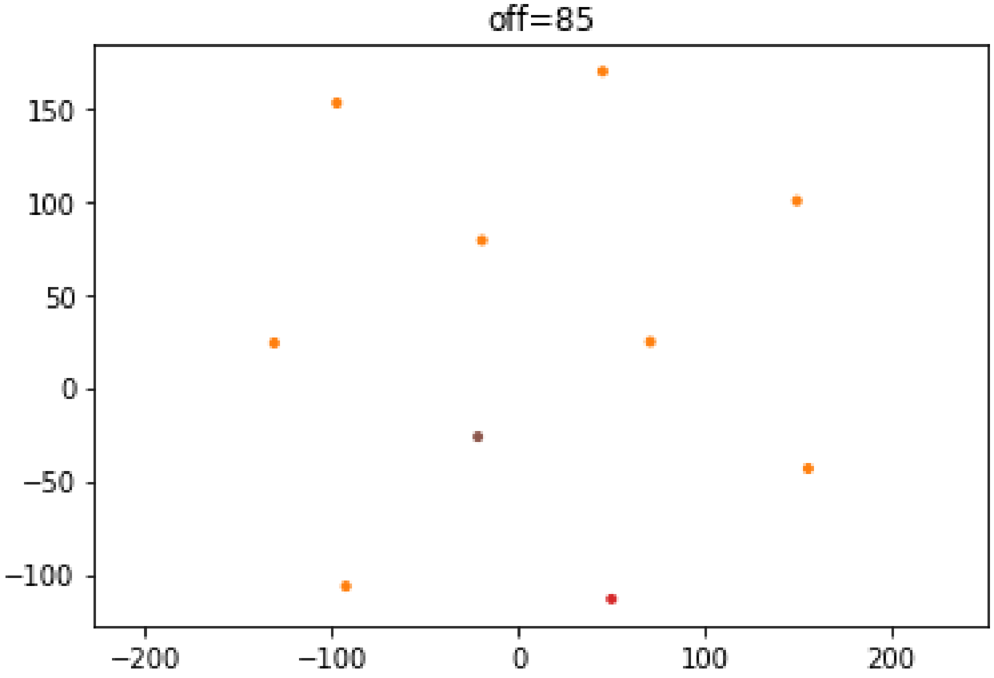

off=80
300
200
100
0
−100
−200
−200
0
200
400

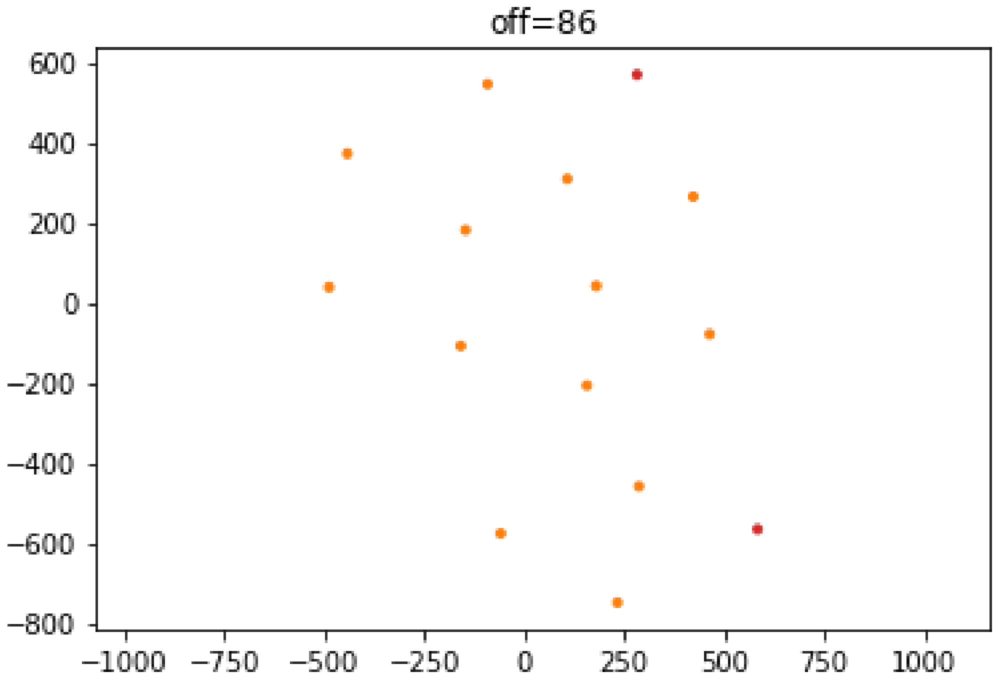

off=81
150
100
50
0
−50
−400
−350
−300
−250
−200
−150
−100
−50

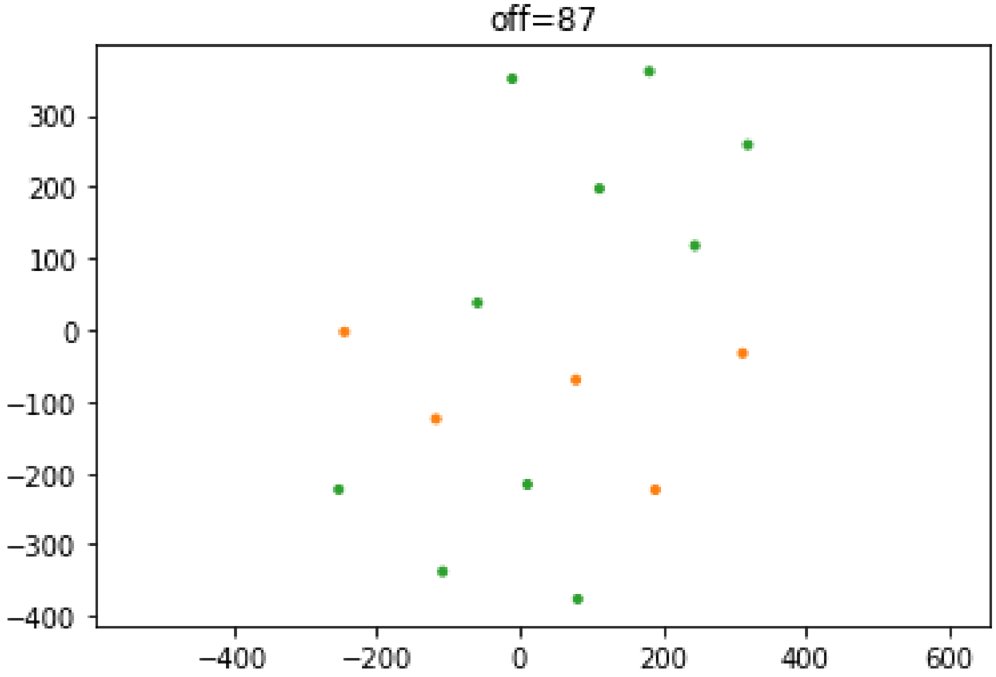

off=82
100
50
0
−50
−100
−150
−100
−50
0
50
100
150

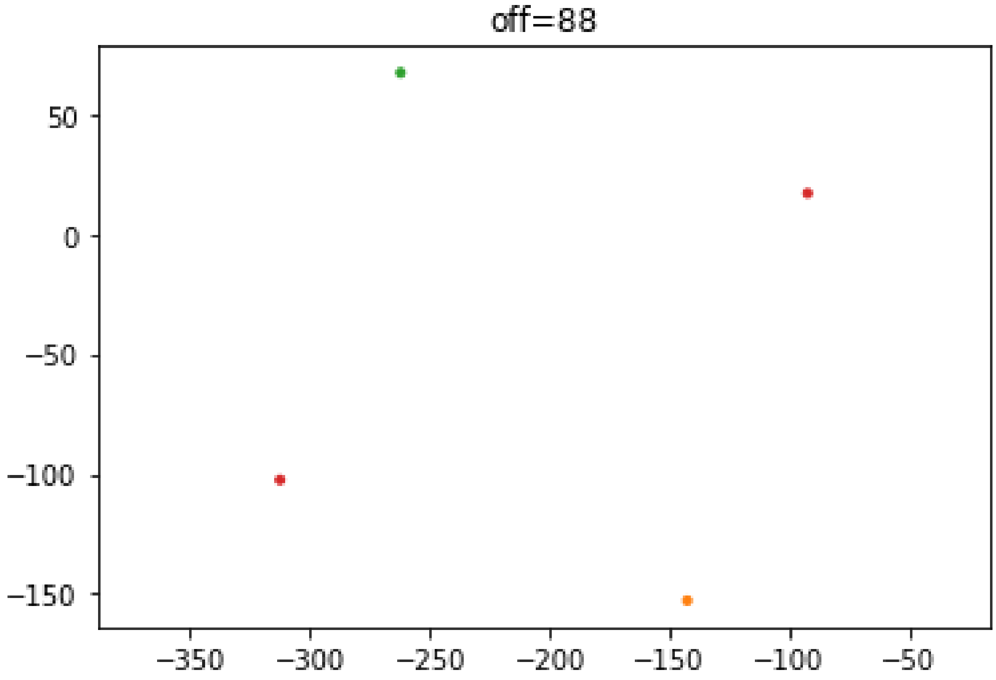

off=83
60
40
20
0
−20
−40
−60
−125
−100
−75
−50
−25
0
25
50

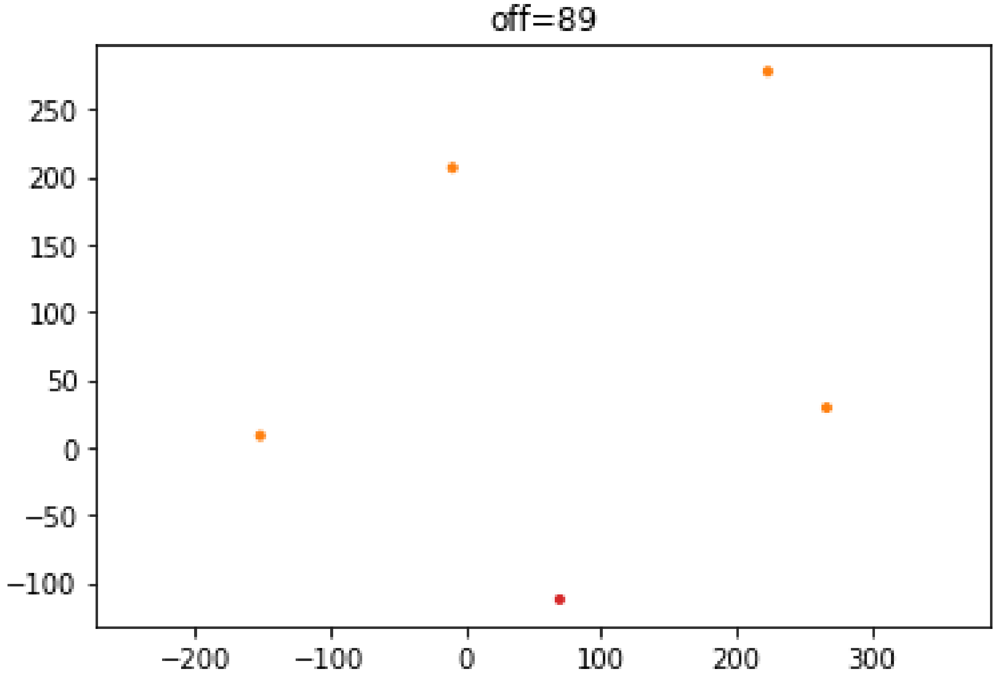

off=84
150
100
50
0
−50
−100
−150
−200
−100
0
100
200
300

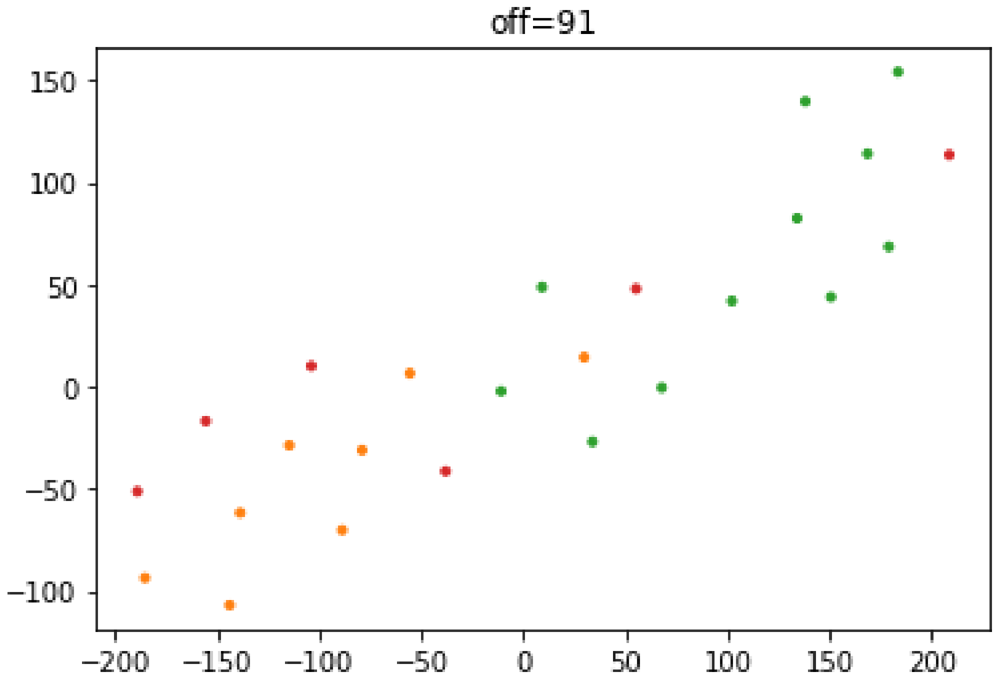

off=85
150
100
50
0
−50
−100
−200
−100
0
100
200

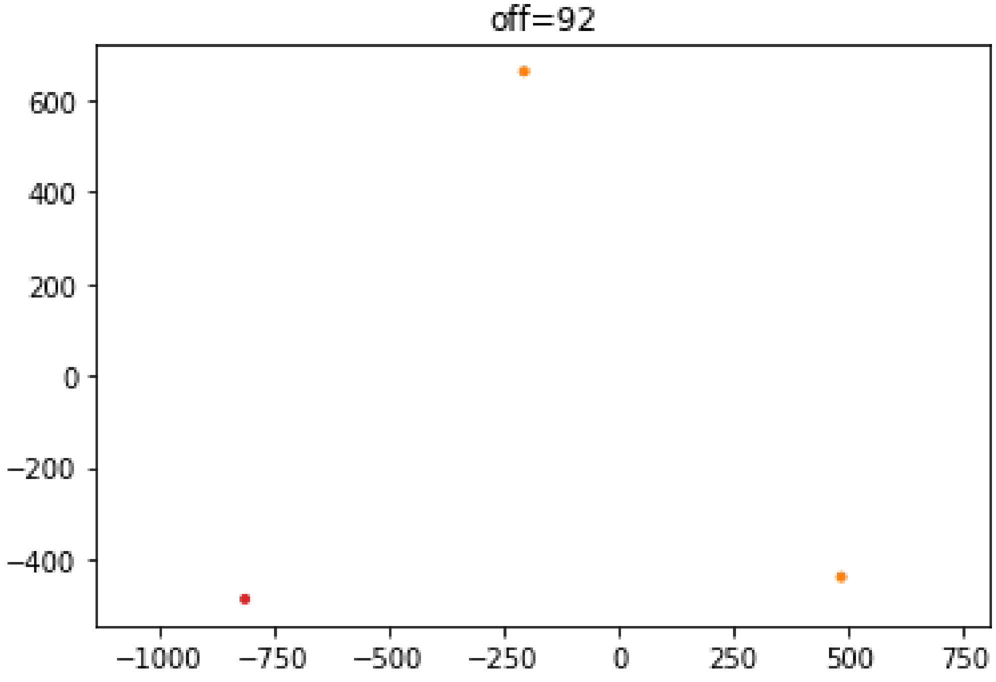

off=86
600
400
200
0
−200
−400
−600
−800
−1000
−750
−500
−250
0
250
500
750
1000

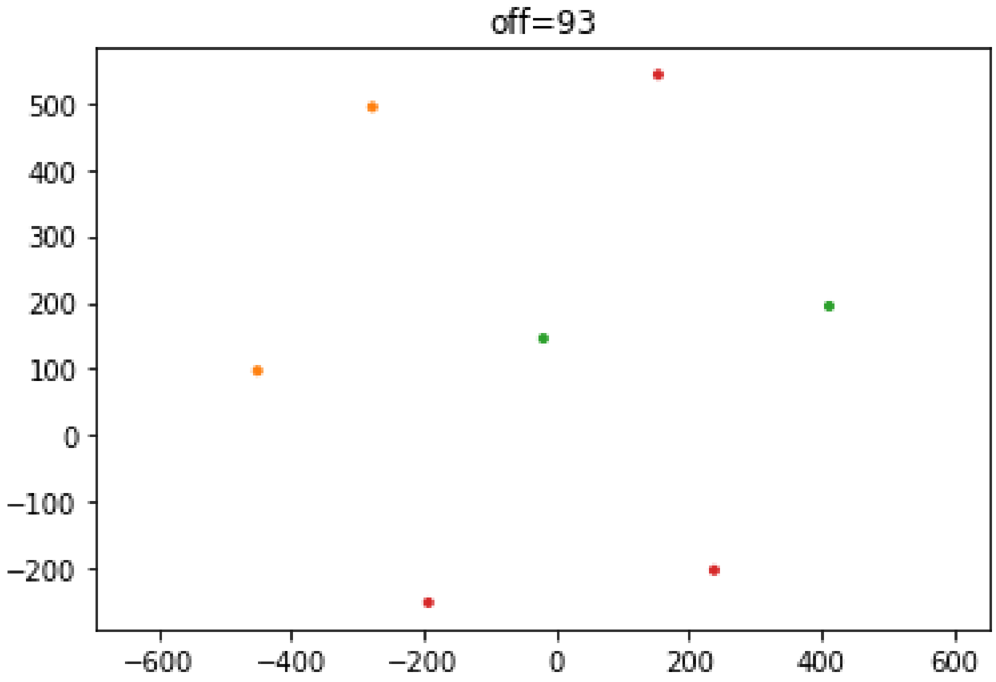

off=87
300
200
100
0
−100
−200
−300
−400
−400
−200
0
200
400
600

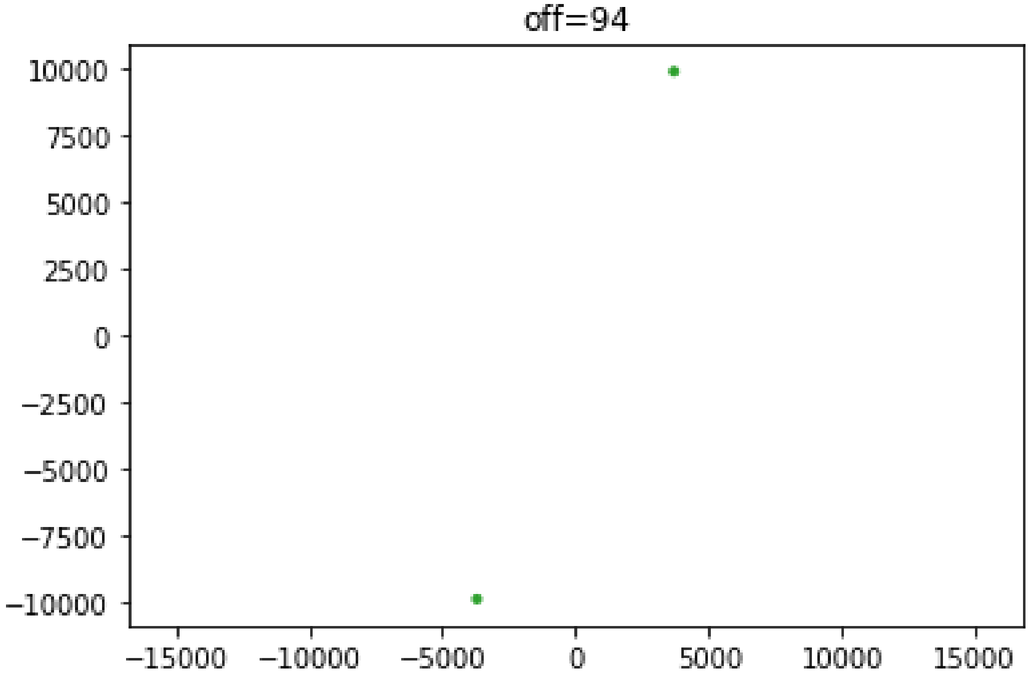

off=88
50
0
−50
−100
−150
−350
−300
−250
−200
−150
−100
−50

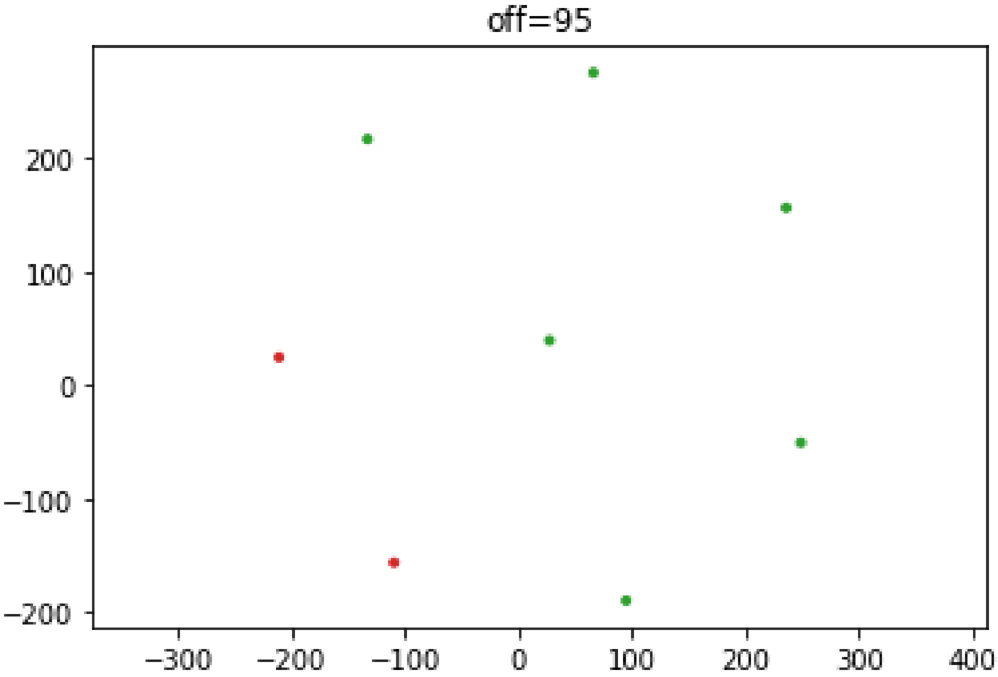

off=89
250
200
150
100
50
0
−50
−100
−200
−100
0
100
200
300

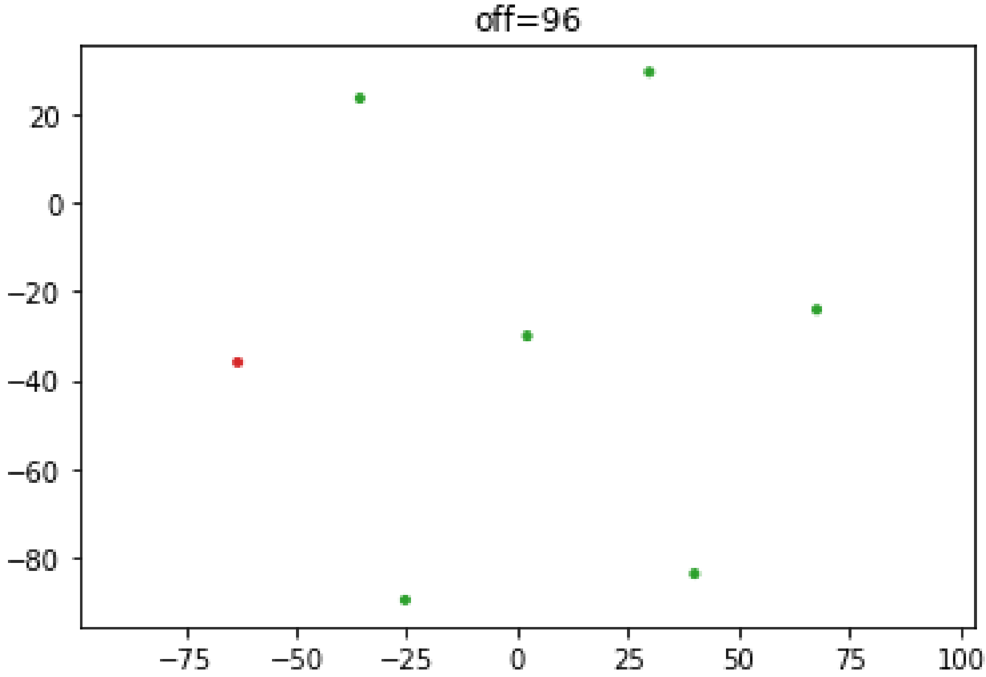

off=91
150
100
50
0
−50
−100
−200
−150
−100
−50
0
50
100
150
200

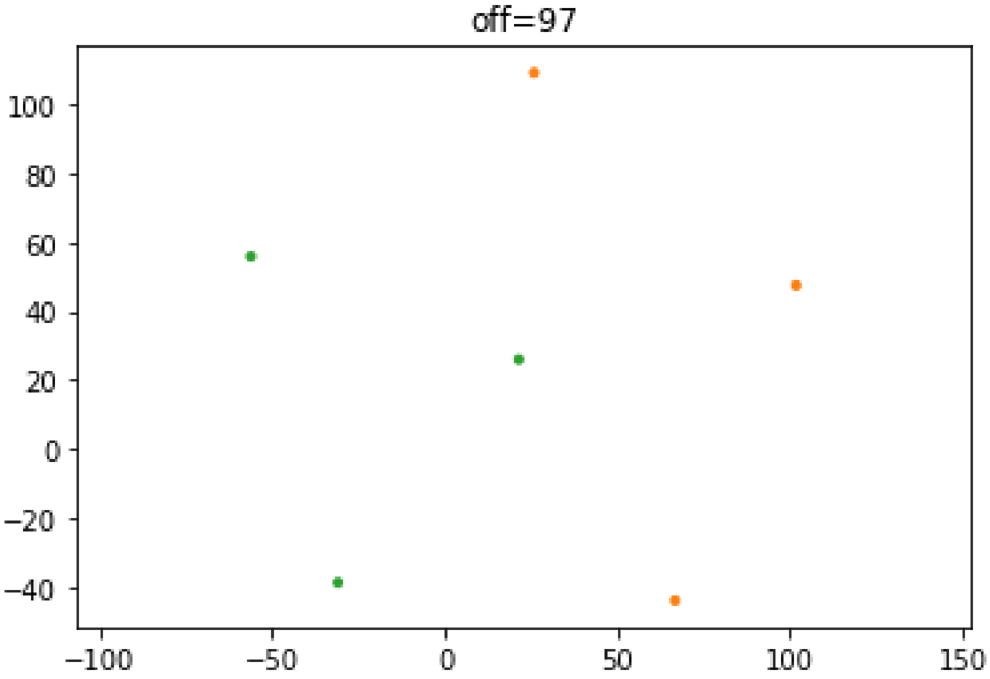

off=92
600
400
200
0
−200
−400
−1000
−750
−500
−250
0
250
500
750

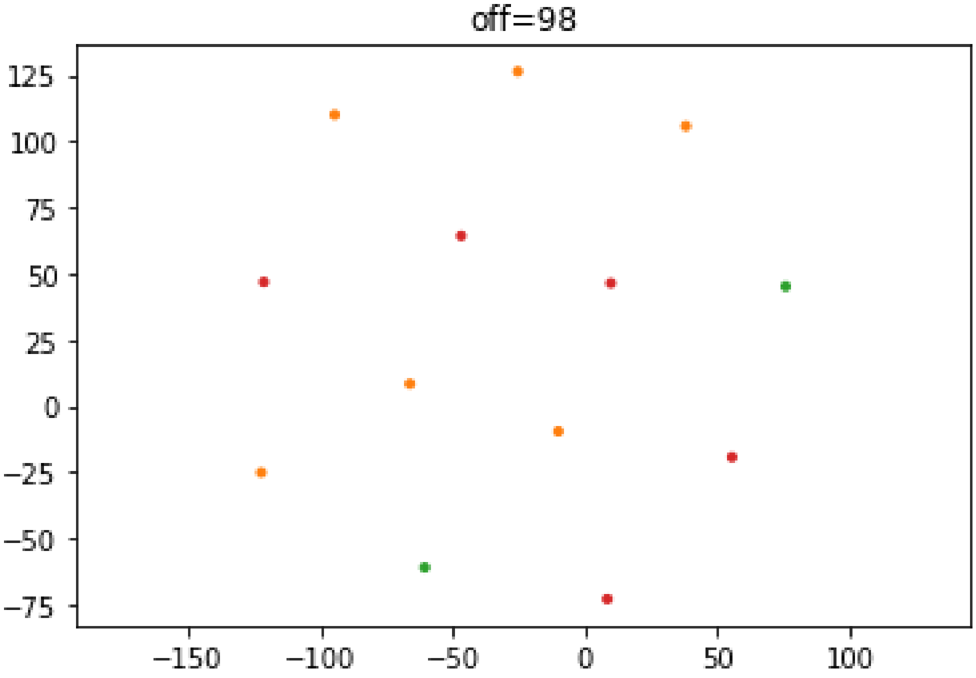

off=93
500
400
300
200
100
0
−100
−200
−600
−400
−200
0
200
400
600

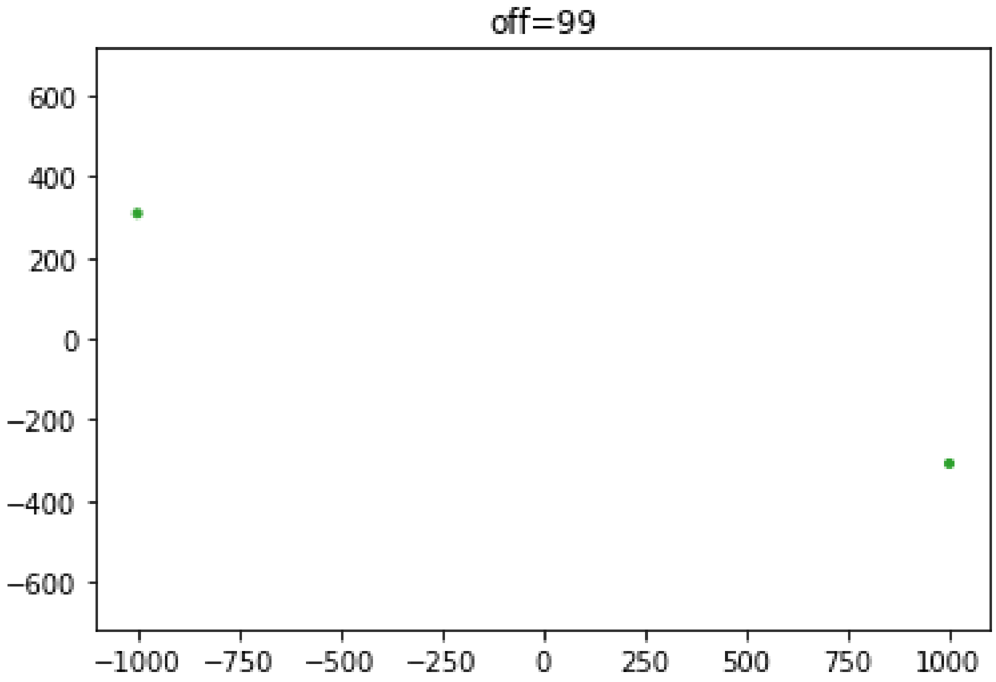

off=94
10000
7500
5000
2500
0
−2500
−5000
−7500
−10000
−15000
−10000
−5000
0
5000
10000
15000

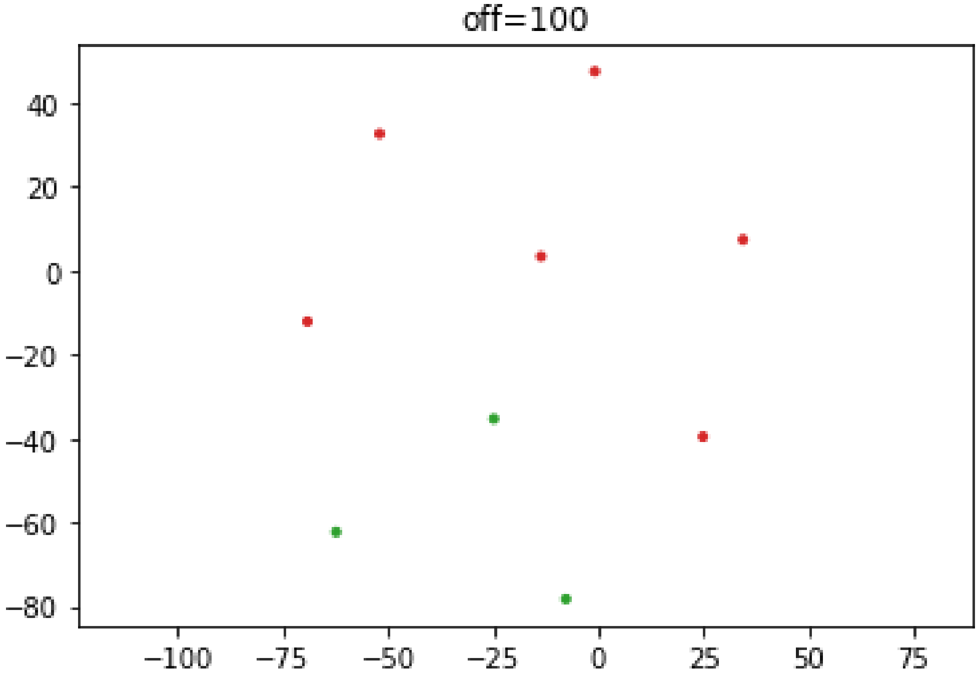

off=95
200
100
0
−100
−200
−300
−200
−100
0
100
200
300
400

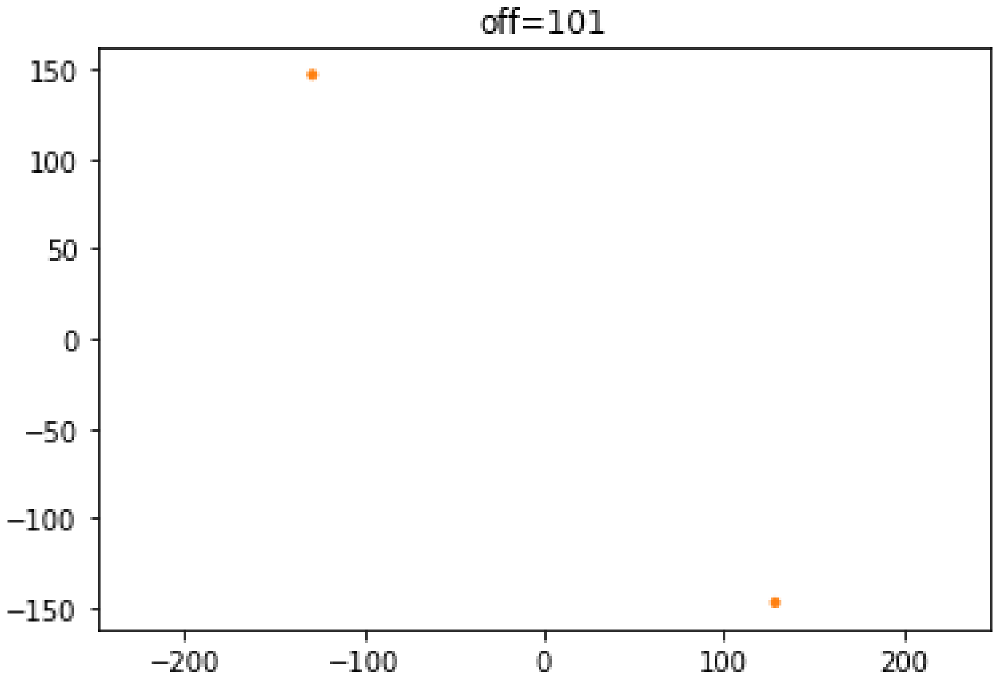

off=96
20
0
−20
−40
−60
−80
−75
−50
−25
0
25
50
75
100

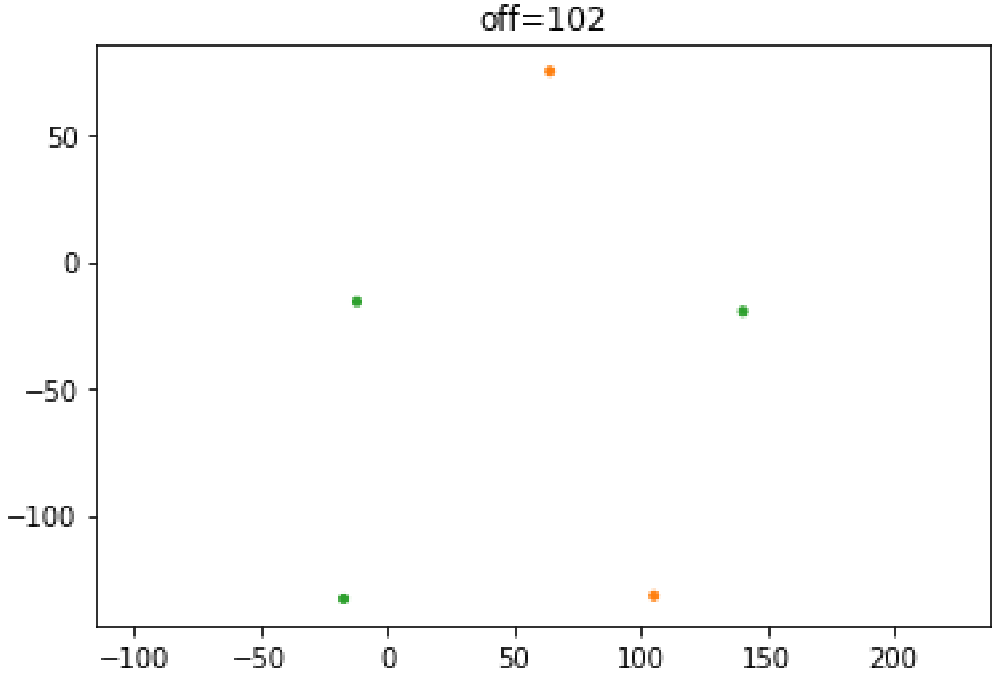

off=97
100
80
60
40
20
0
−20
−40
−100
−50
0
50
100
150

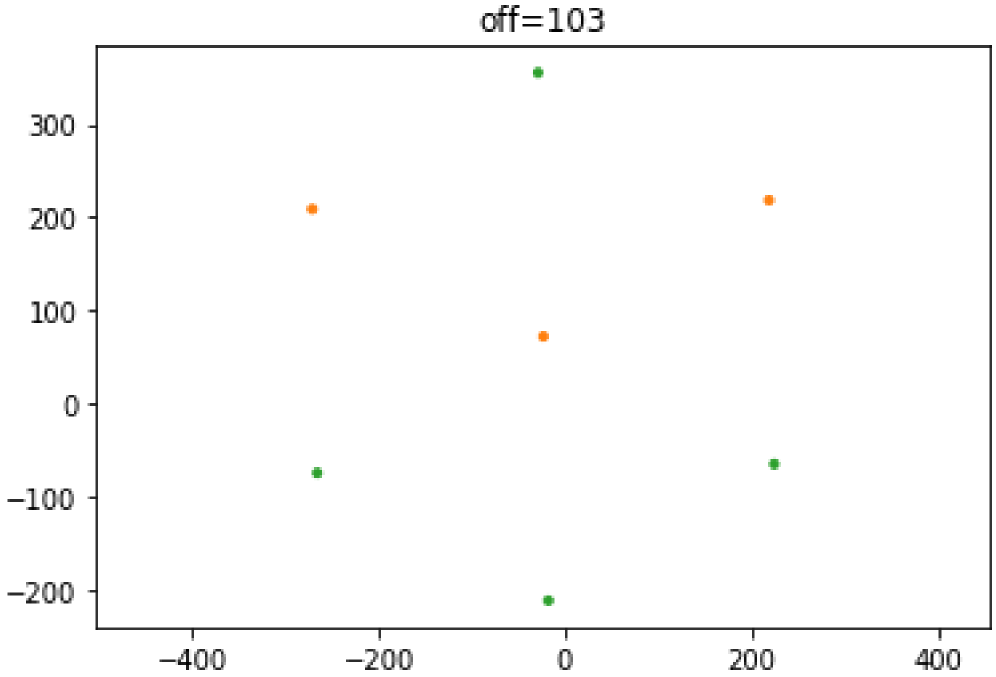

off=98
125
100
75
50
25
0
−25
−50
−75
−150
−100
−50
0
50
100

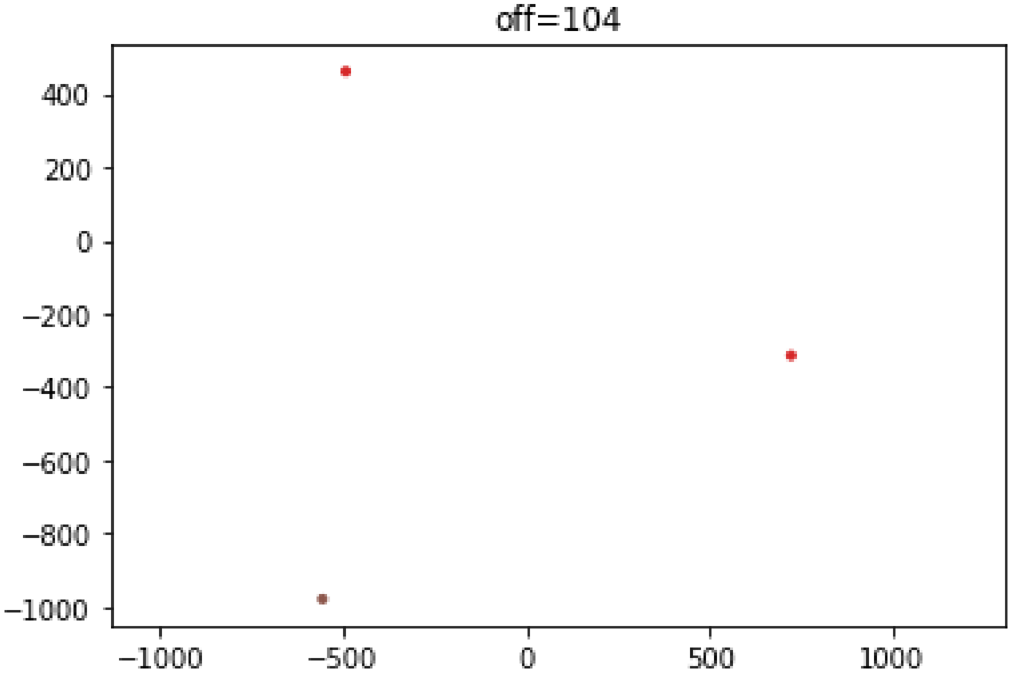

off=99
600
400
200
0
−200
−400
−600
−1000
−750
−500
−250
0
250
500
750
1000

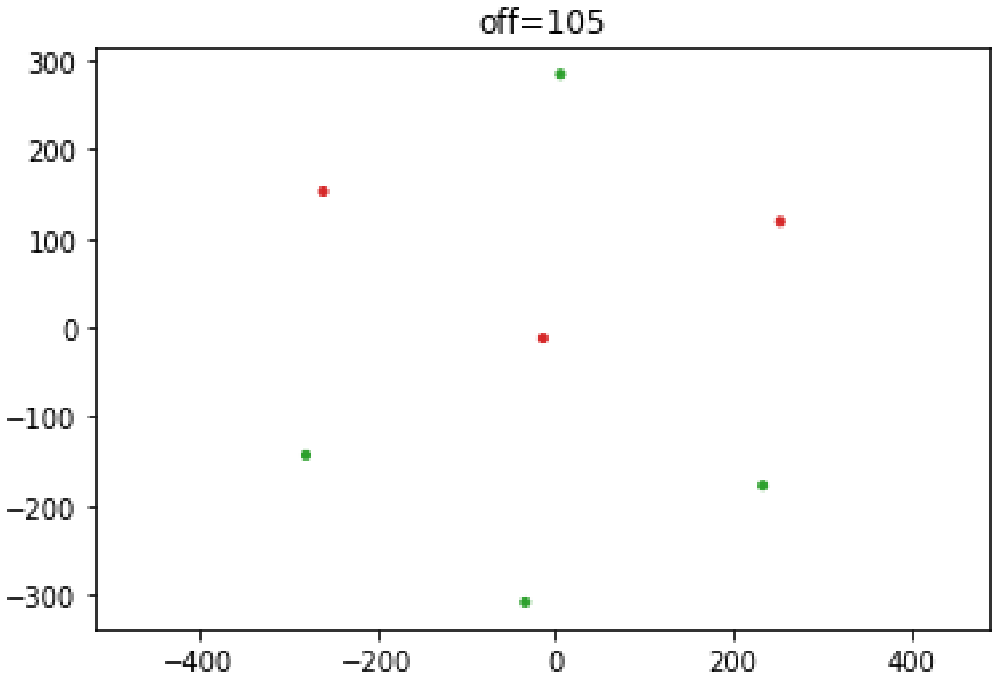

off=100
40
20
0
−20
−40
−60
−80
−100
−75
−50
−25
0
25
50
75

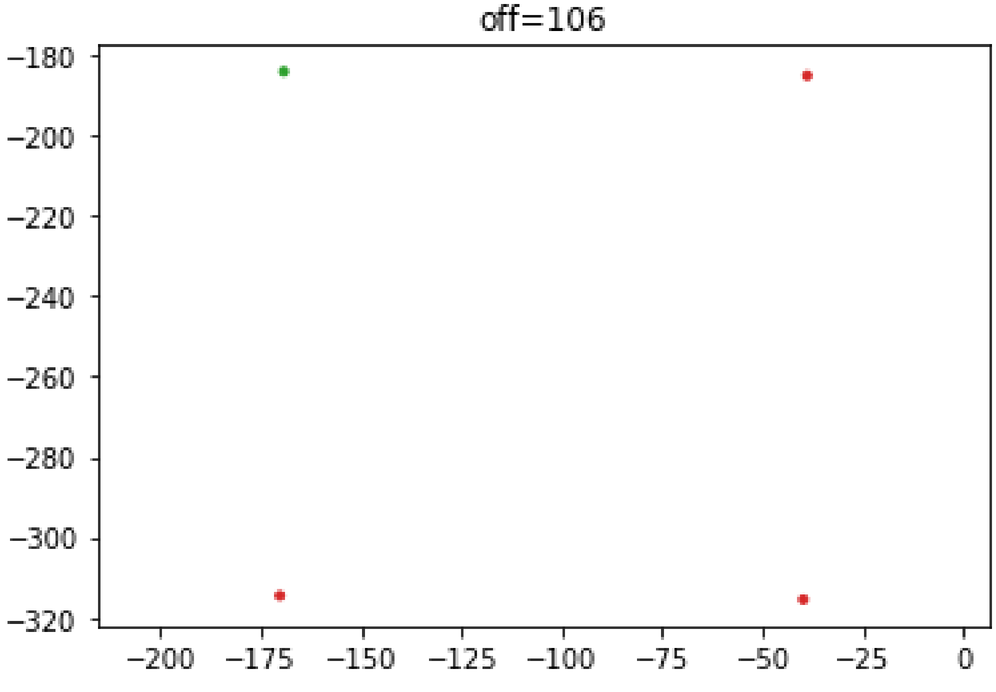

off=101
150
100
50
0
−50
−100
−150
−200
−100
0
100
200

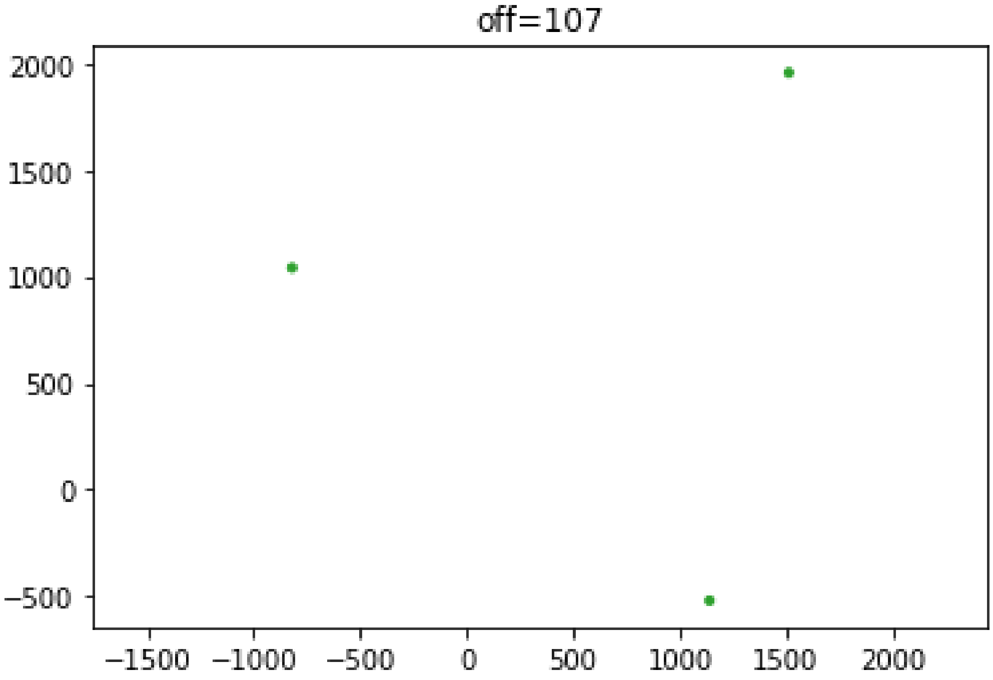

off=102
50
0
−50
−100
−100
−50
0
50
100
150
200

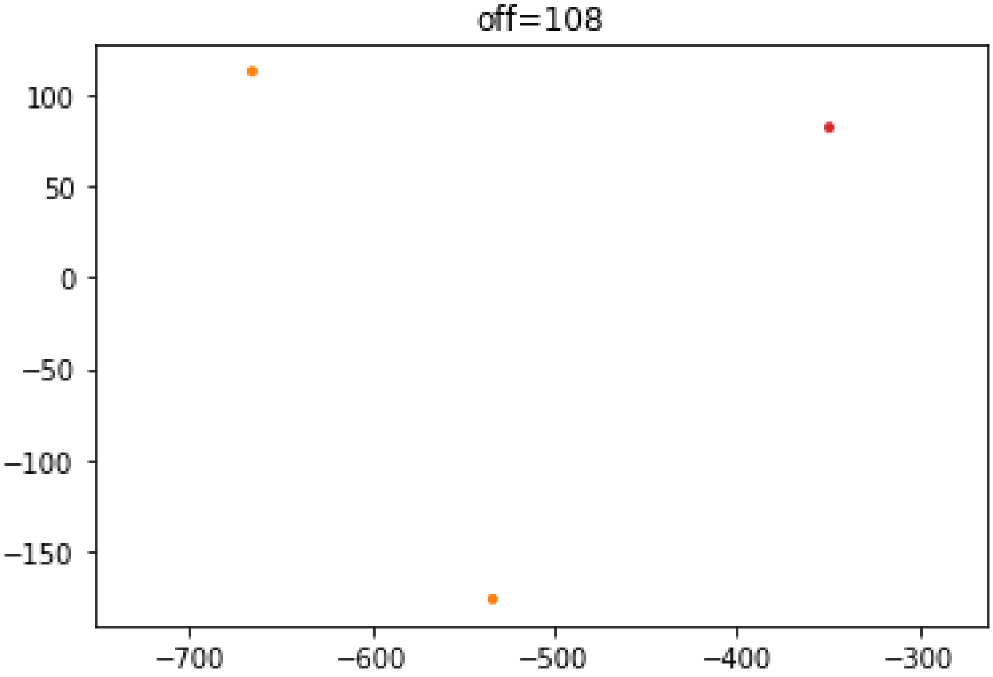

off=103
300
200
100
0
−100
−200
−400
−200
0
200
400

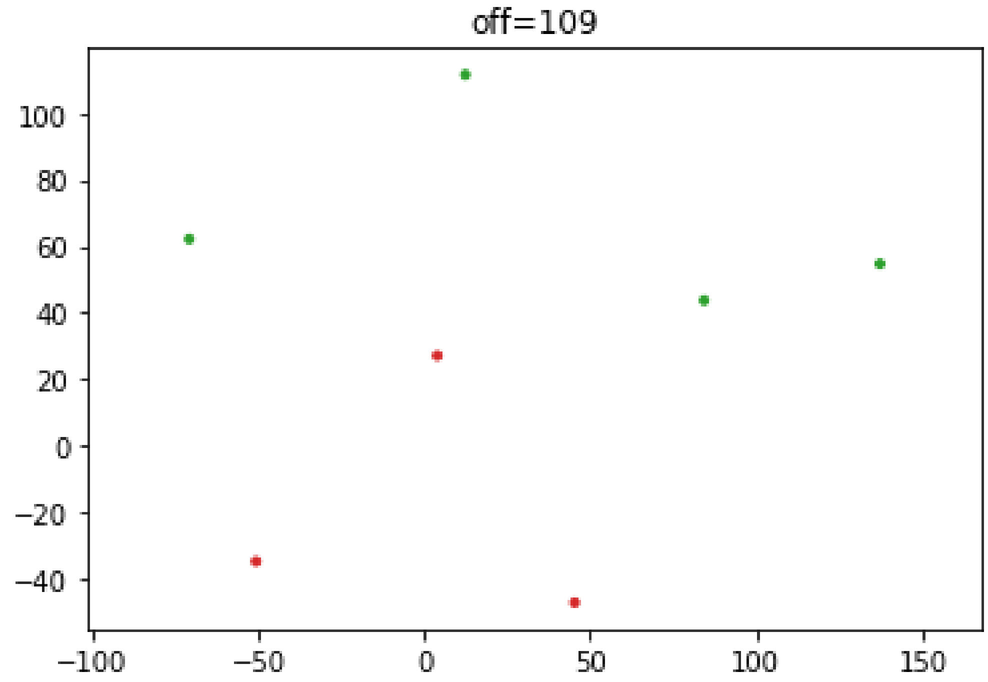

off=104
400
200
0
−200
−400
−600
−800
−1000
−1000
−500
0
500
1000

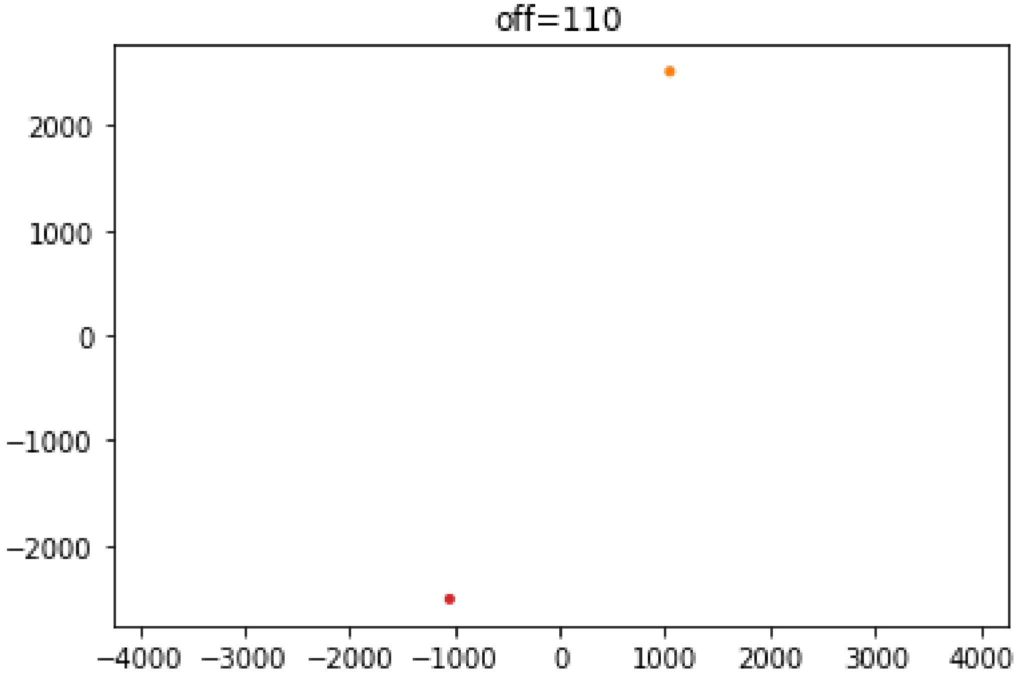

off=105
300
200
100
0
−100
−200
−300
−400
−200
0
200
400

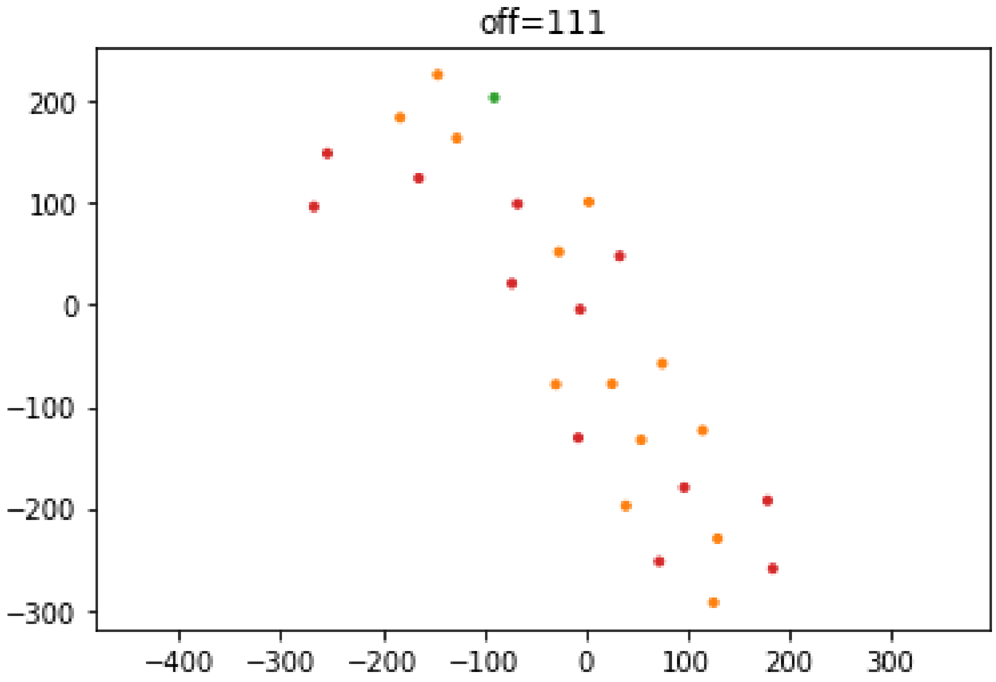

off=106
−180
−200
−220
−240
−260
−280
−300
−320
−200
−175
−150
−125
−100
−75
−50
−25
0

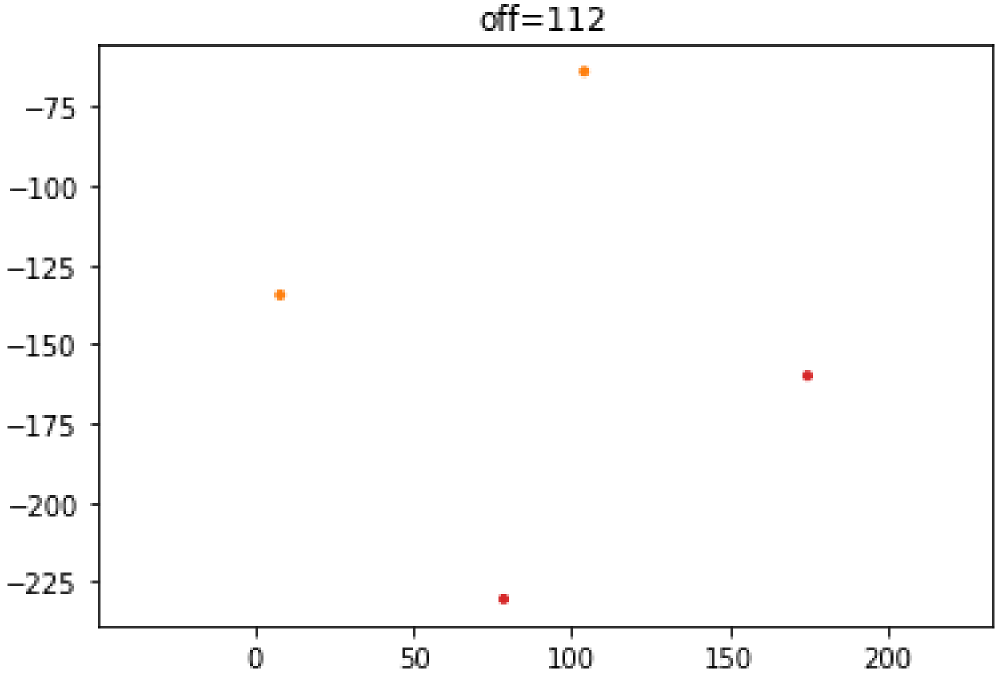

off=107
2000
1500
1000
500
0
−500
−1500
−1000
−500
0
500
1000
1500
2000

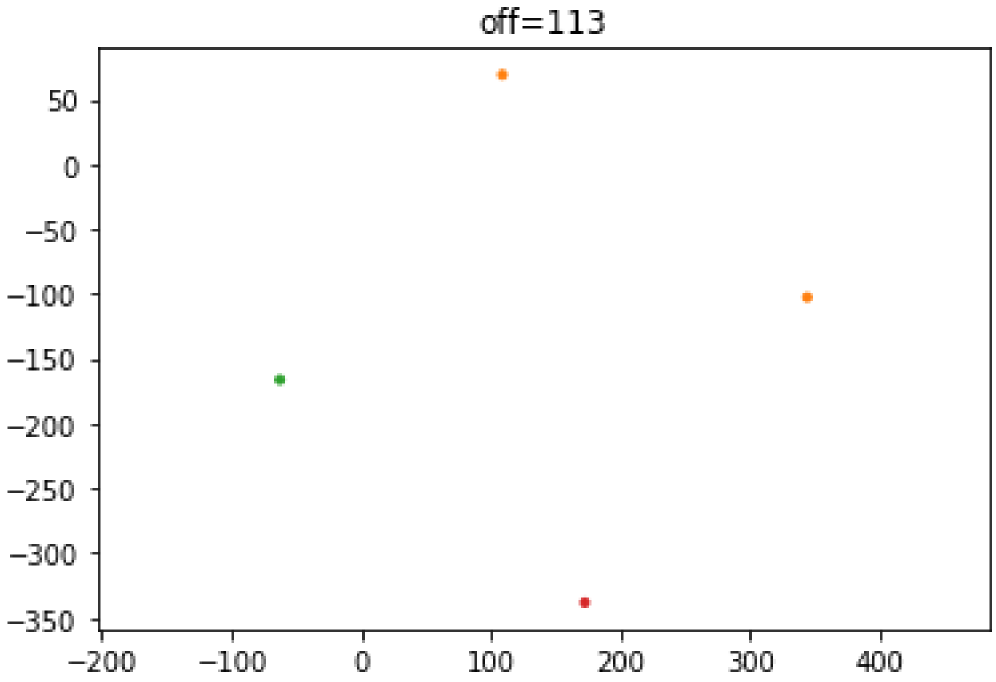

off=108
100
50
0
−50
−100
−150
−700
−600
−500
−400
−300

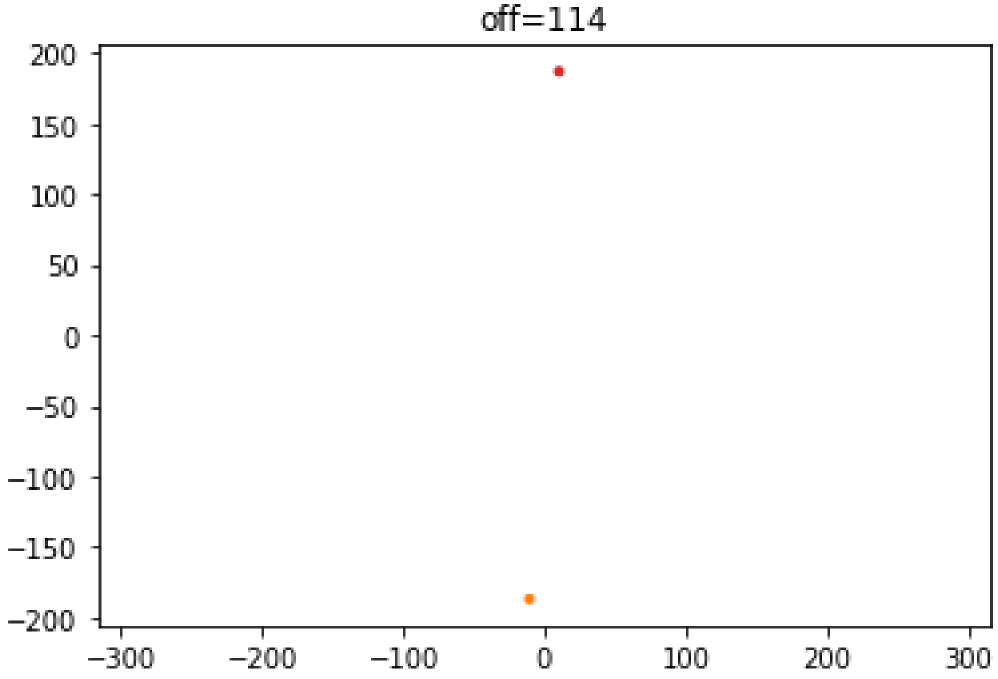

off=109
100
80
60
40
20
0
−20
−40
−100
−50
0
50
100
150

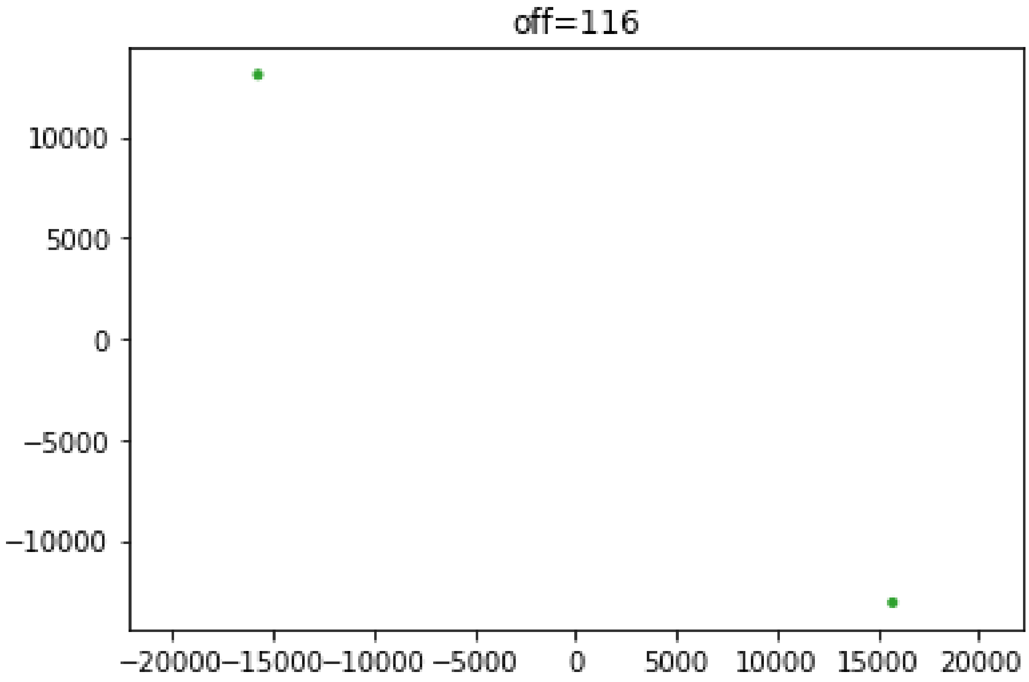

off=110
2000
1000
0
−1000
−2000
−4000
−3000
−2000
−1000
0
1000
2000
3000
4000

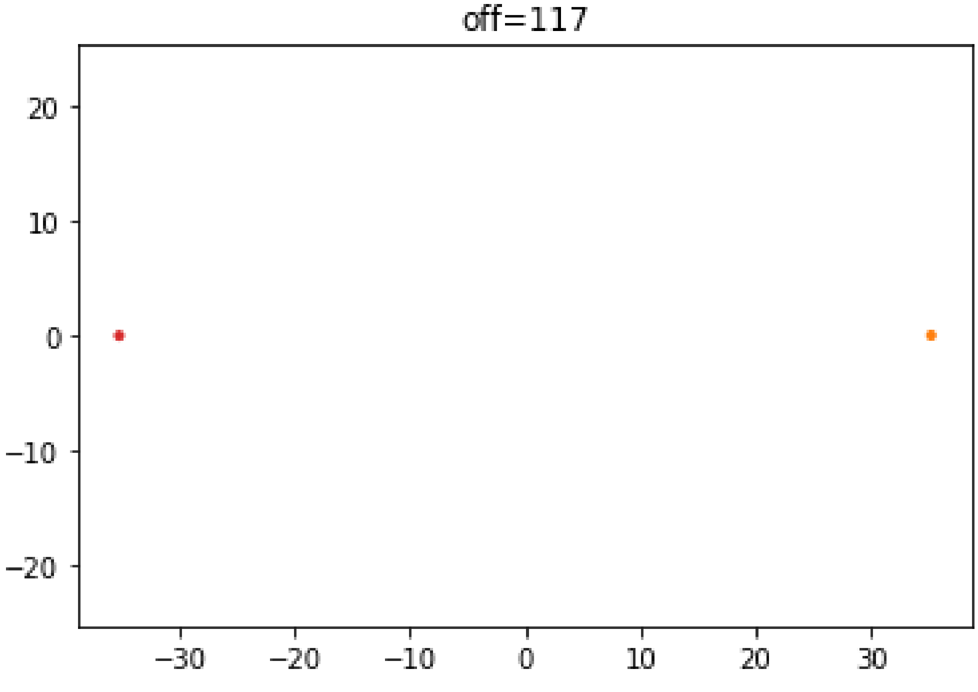

off=111
200
100
0
−100
−200
−300
−400
−300
−200
−100
0
100
200
300

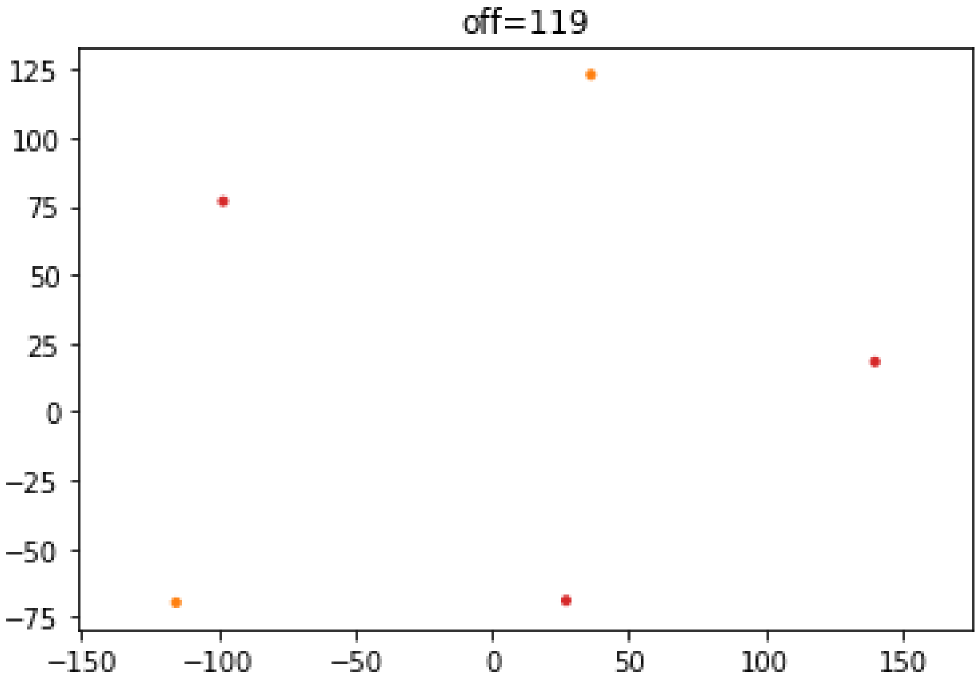

off=112
−75
−100
−125
−150
−175
−200
−225
0
50
100
150
200

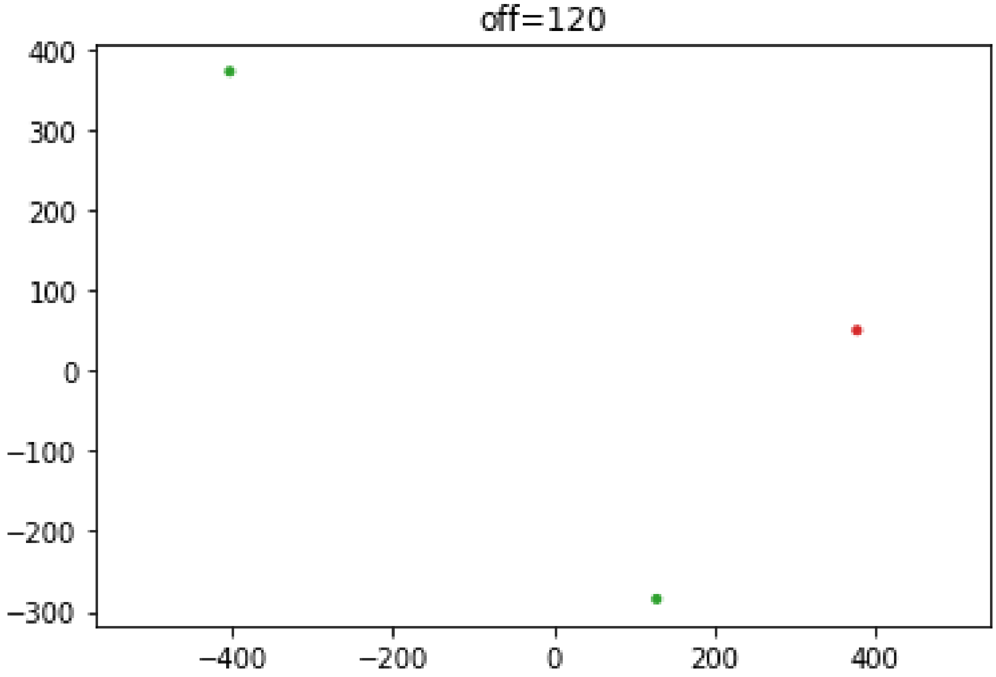

off=113
50
0
−50
−100
−150
−200
−250
−300
−350
−200
−100
0
100
200
300
400

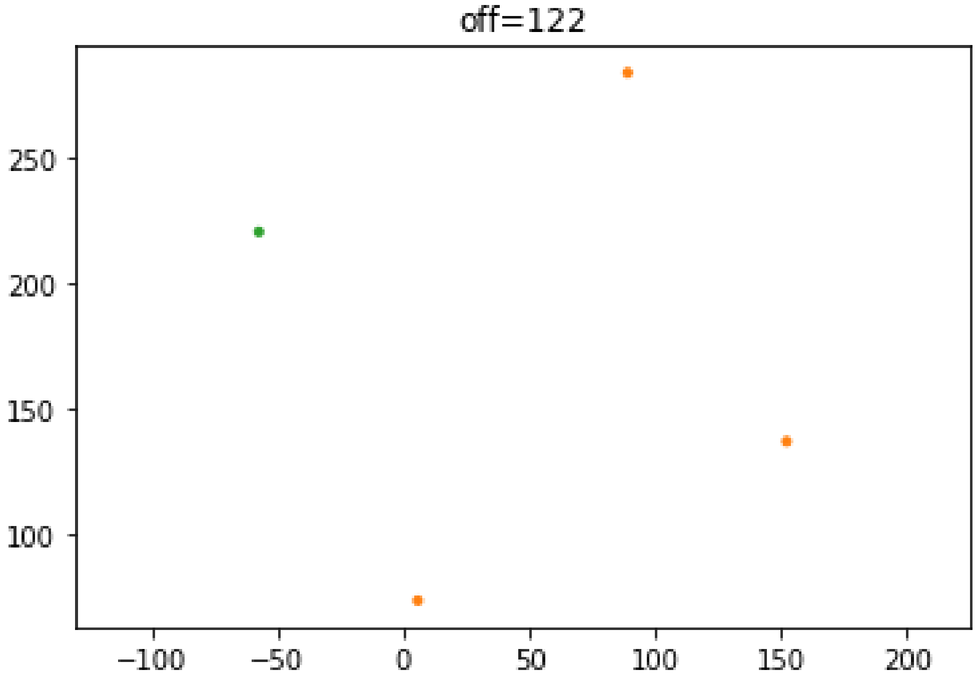

off=114
200
150
100
50
0
−50
−100
−150
−200
−300
−200
−100
0
100
200
300

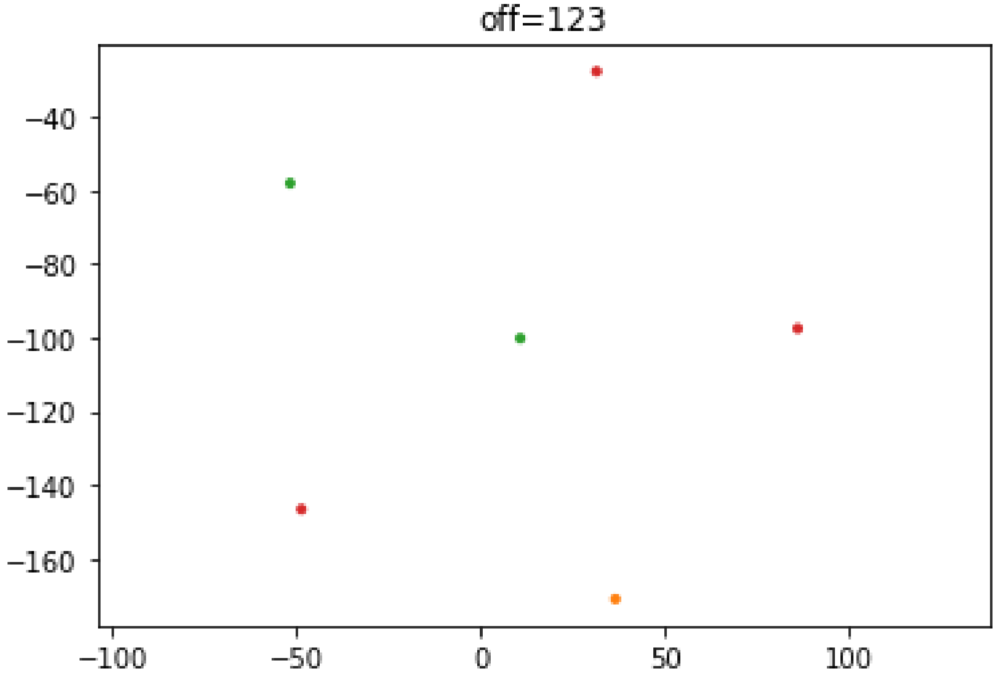

off=116
10000
5000
0
−5000
−10000
−20000
−15000
−10000
−5000
0
5000
10000
15000
20000

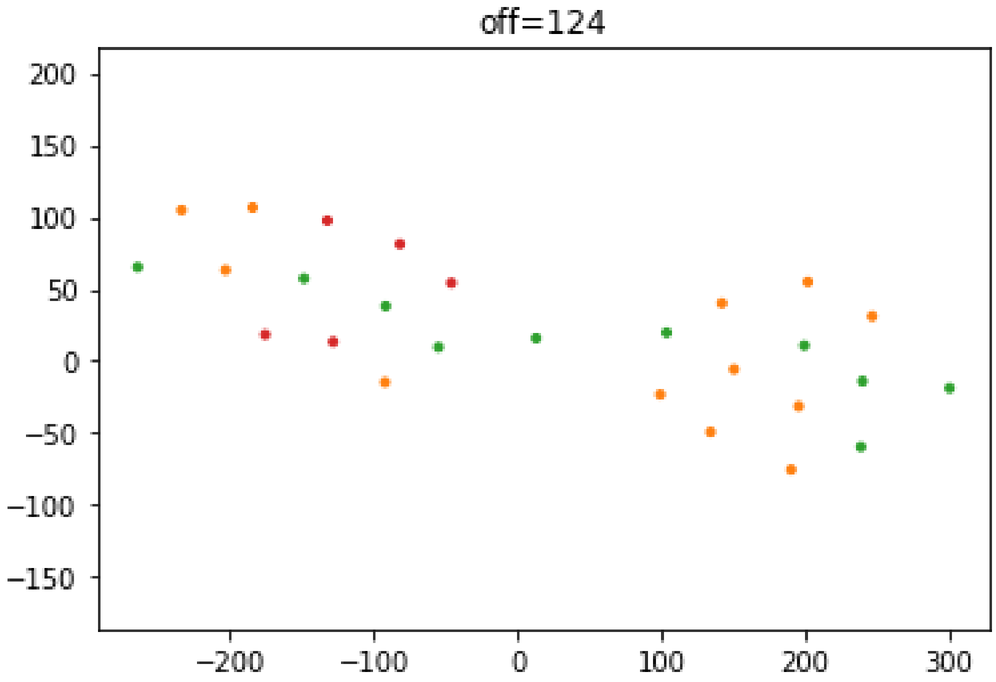

off=117
20
10
0
−10
−20
−30
−20
−10
0
10
20
30

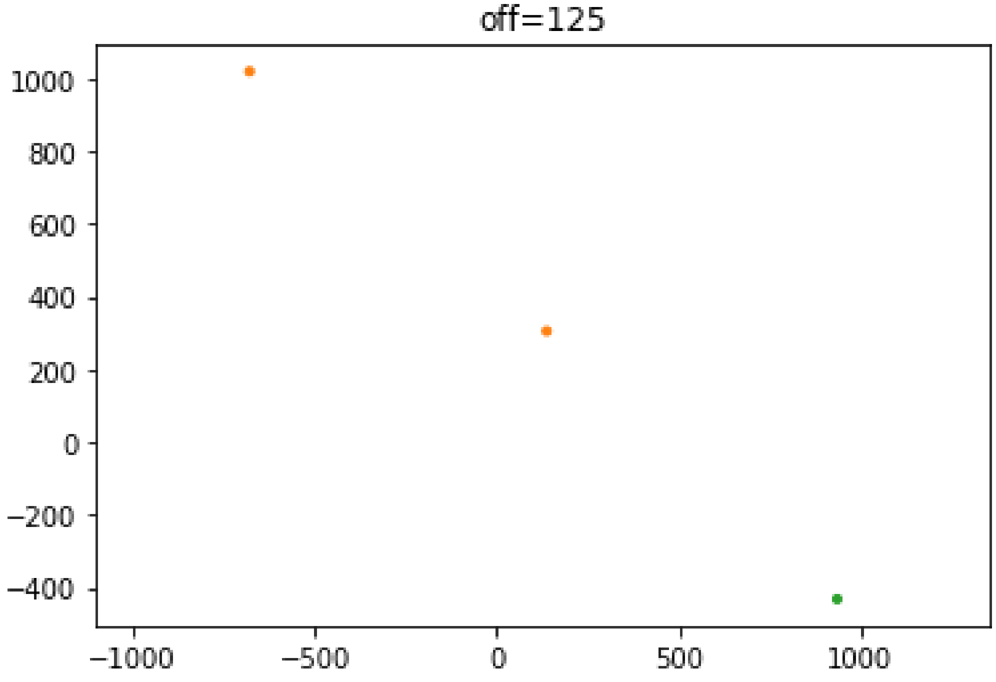

off=119
125
100
75
50
25
0
−25
−50
−75
−150
−100
−50
0
50
100
150

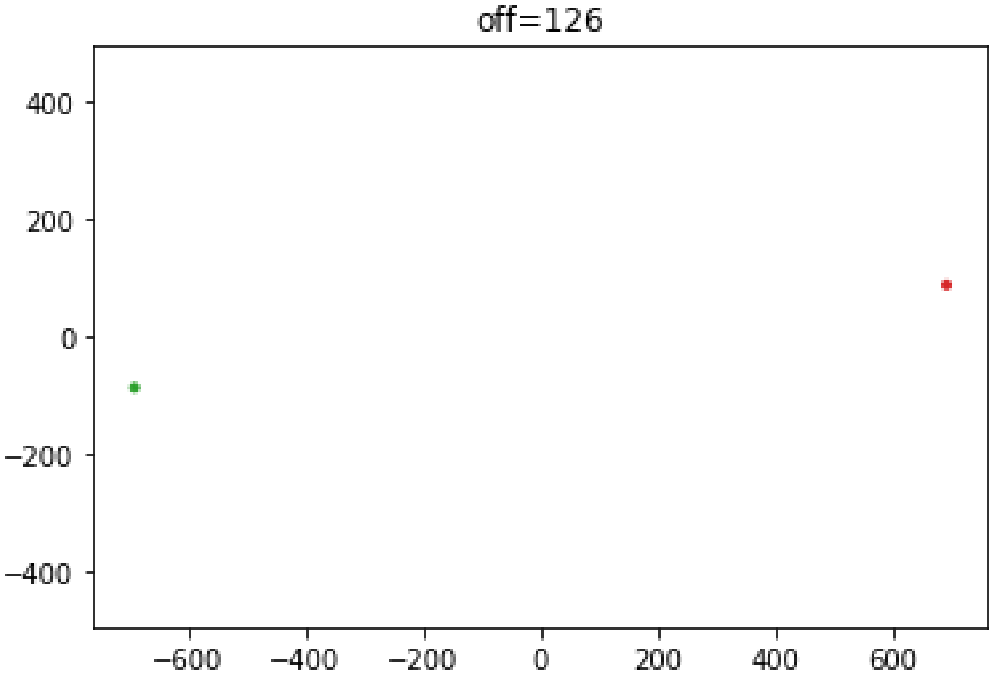

off=120
400
300
200
100
0
−100
−200
−300
−400
−200
0
200
400

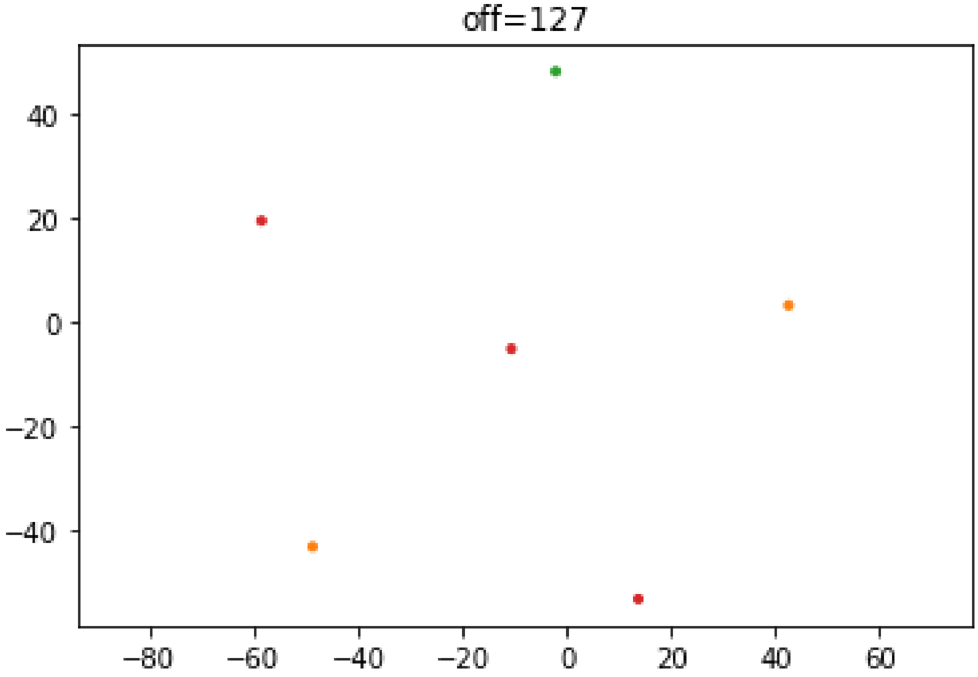

off=122
250
200
150
100
−100
−50
0
50
100
150
200

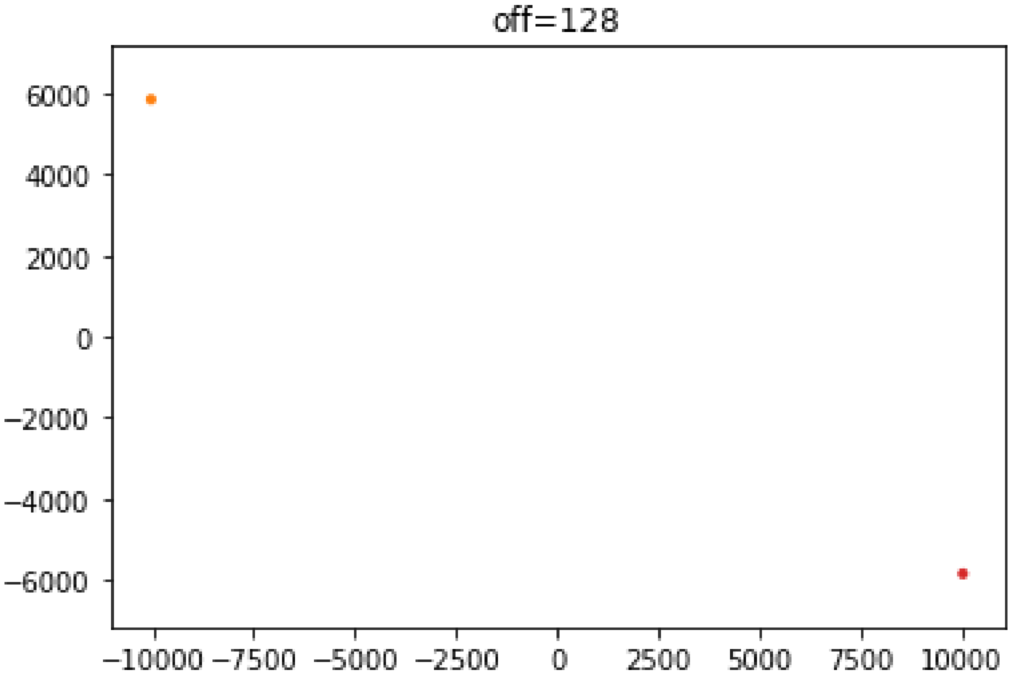

off=123
−40
−60
−80
−100
−120
−140
−160
−100
−50
0
50
100

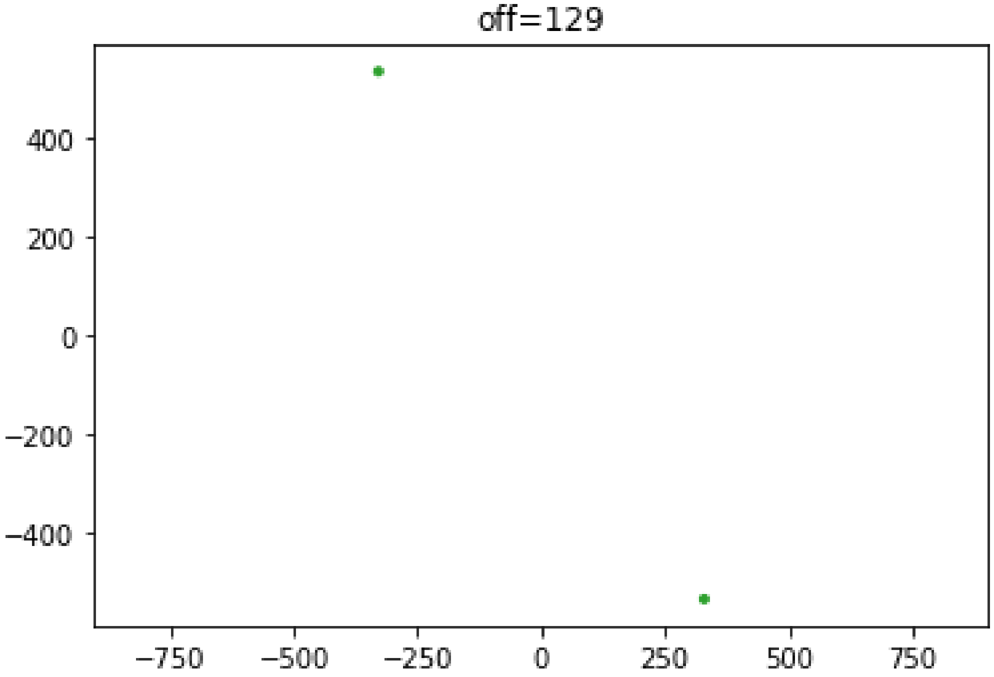

off=124
200
150
100
50
0
−50
−100
−150
−200
−100
0
100
200
300

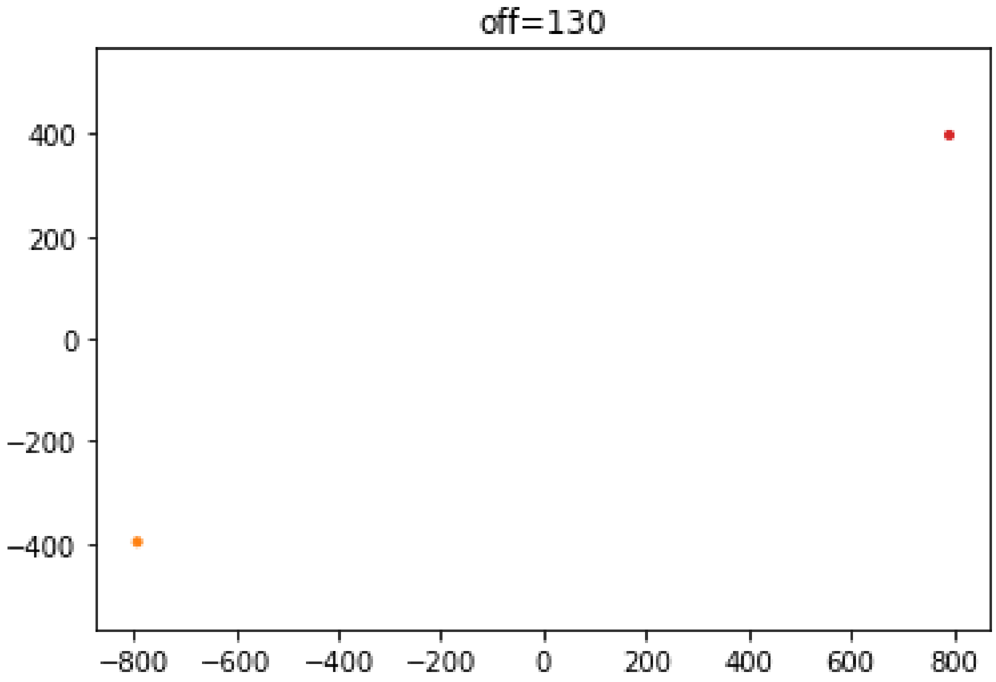

off=125
1000
800
600
400
200
0
−200
−400
−1000
−500
0
500
1000

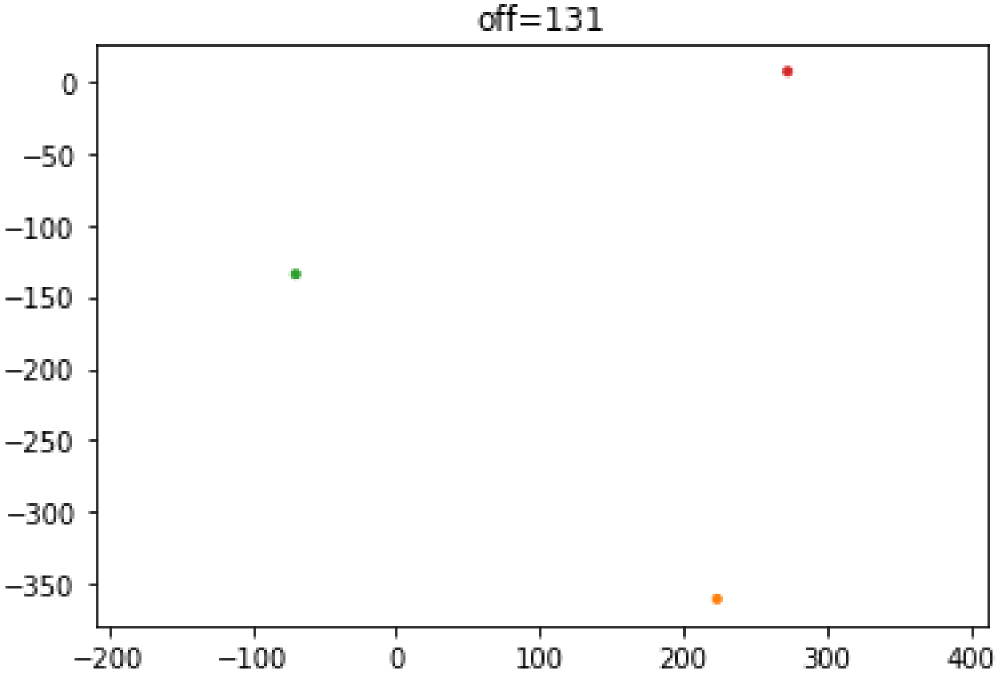

off=126
400
200
0
−200
−400
−600
−400
−200
0
200
400
600

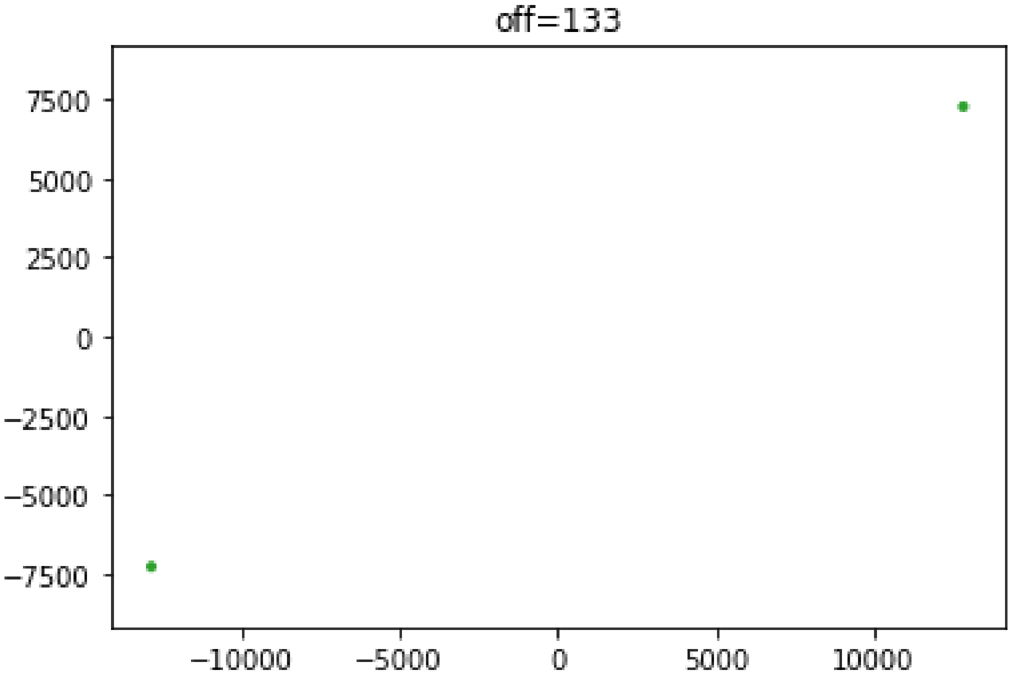

off=127
40
20
0
−20
−40
−80
−60
−40
−20
0
20
40
60

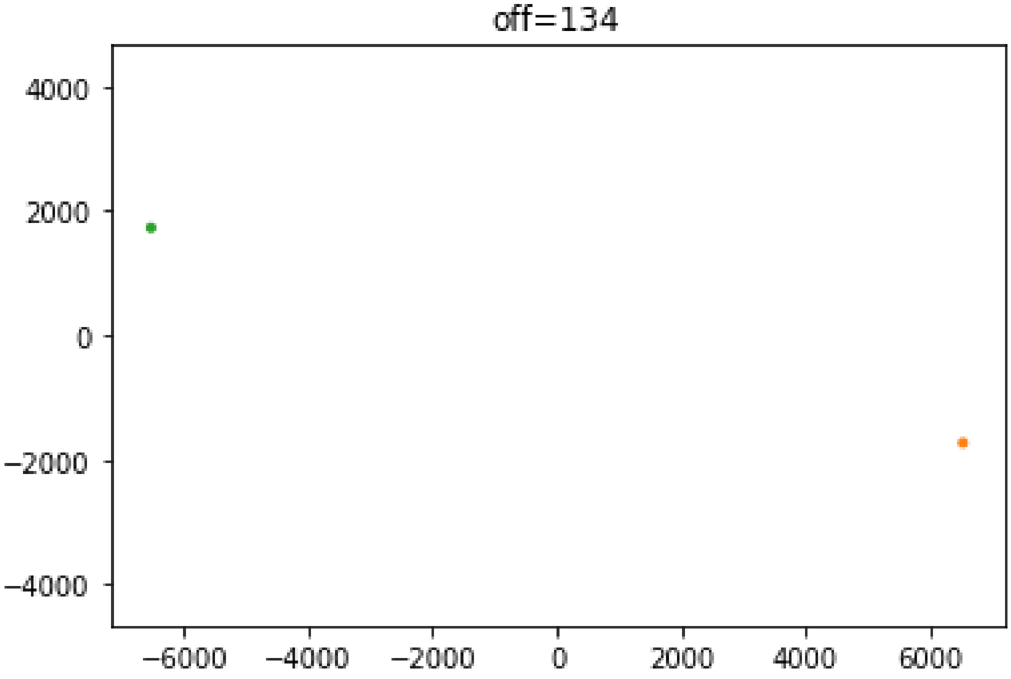

off=128
6000
4000
2000
0
−2000
−4000
−6000
−10000
−7500
−5000
−2500
0
2500
5000
7500
10000

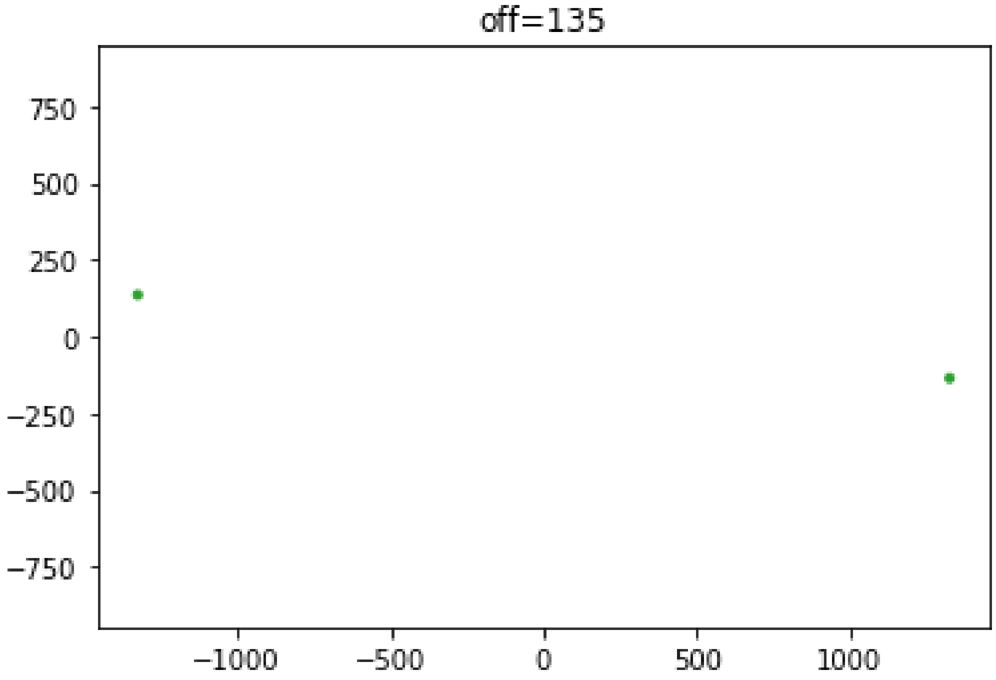

off=129
400
200
0
−200
−400
−750
−500
−250
0
250
500
750

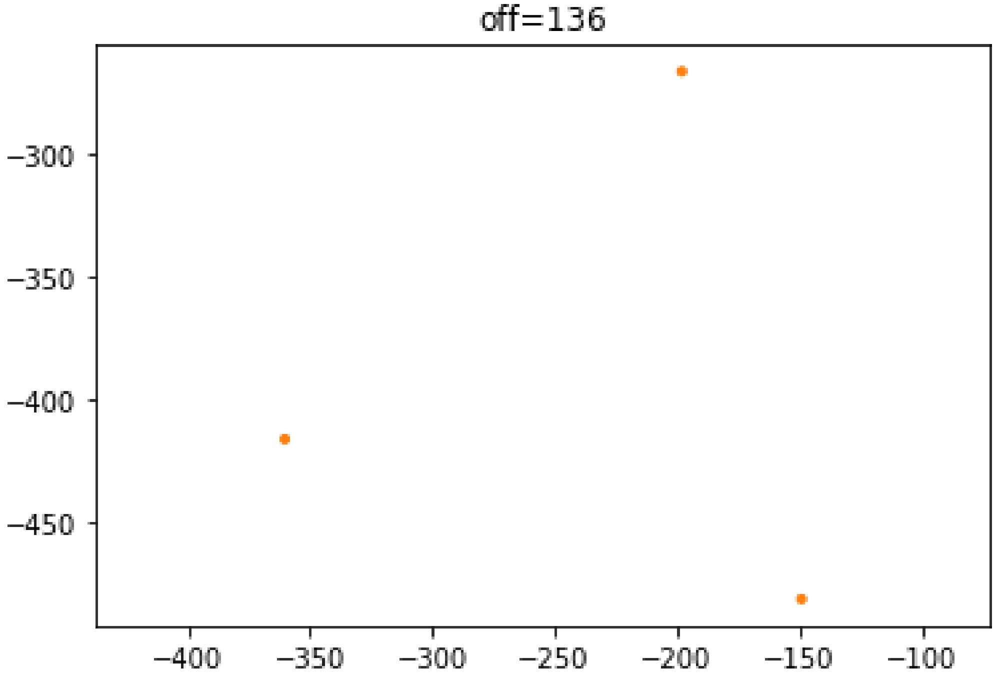

off=130
400
200
0
−200
−400
−800
−600
−400
−200
0
200
400
600
800

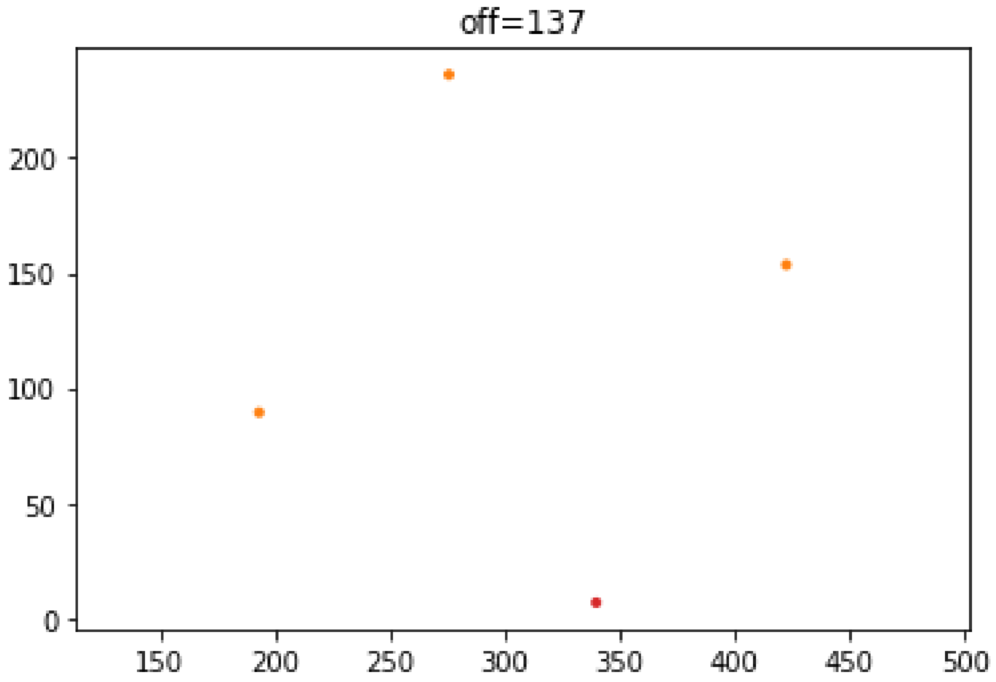

off=131
0
−50
−100
−150
−200
−250
−300
−350
−200
−100
0
100
200
300
400

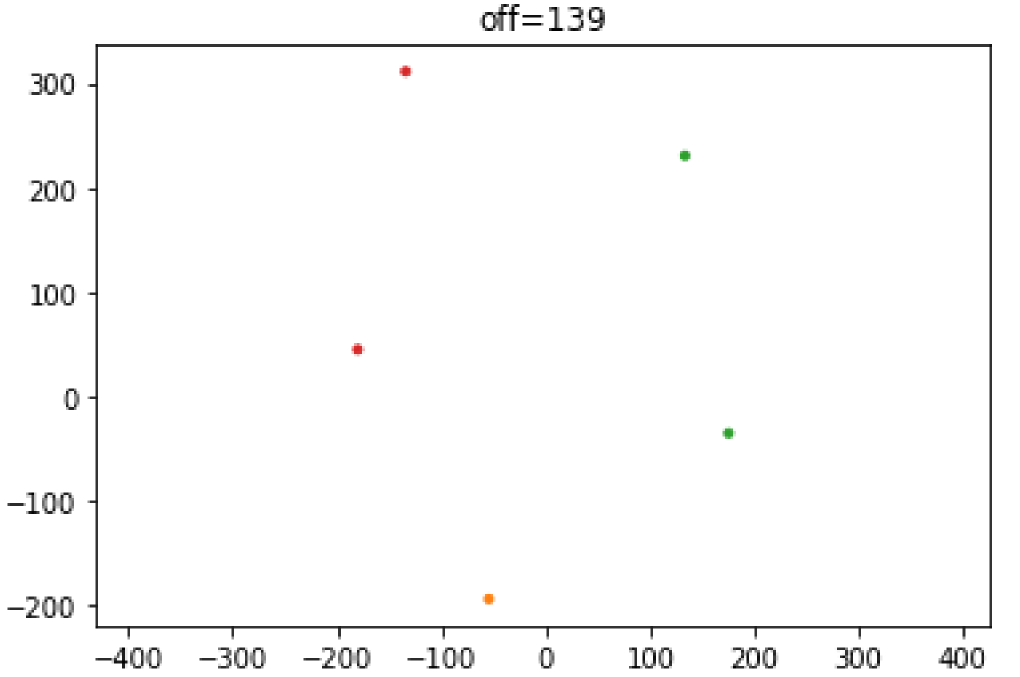

off=133
7500
5000
2500
0
−2500
−5000
−7500
−10000
−5000
0
5000
10000

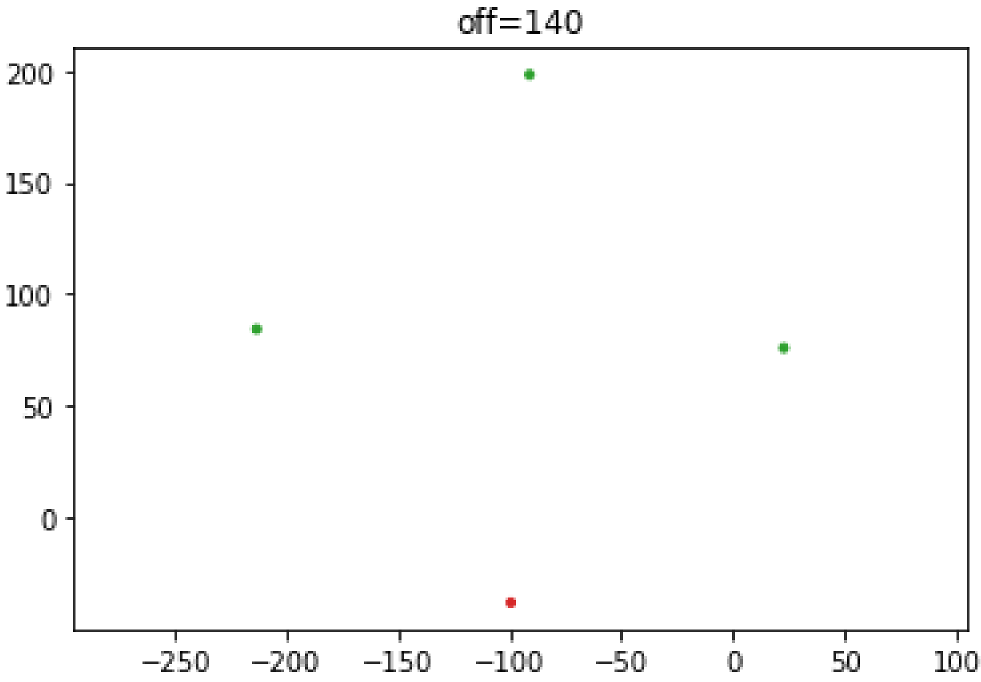

off=134
4000
2000
0
−2000
−4000
−6000
−4000
−2000
0
2000
4000
6000

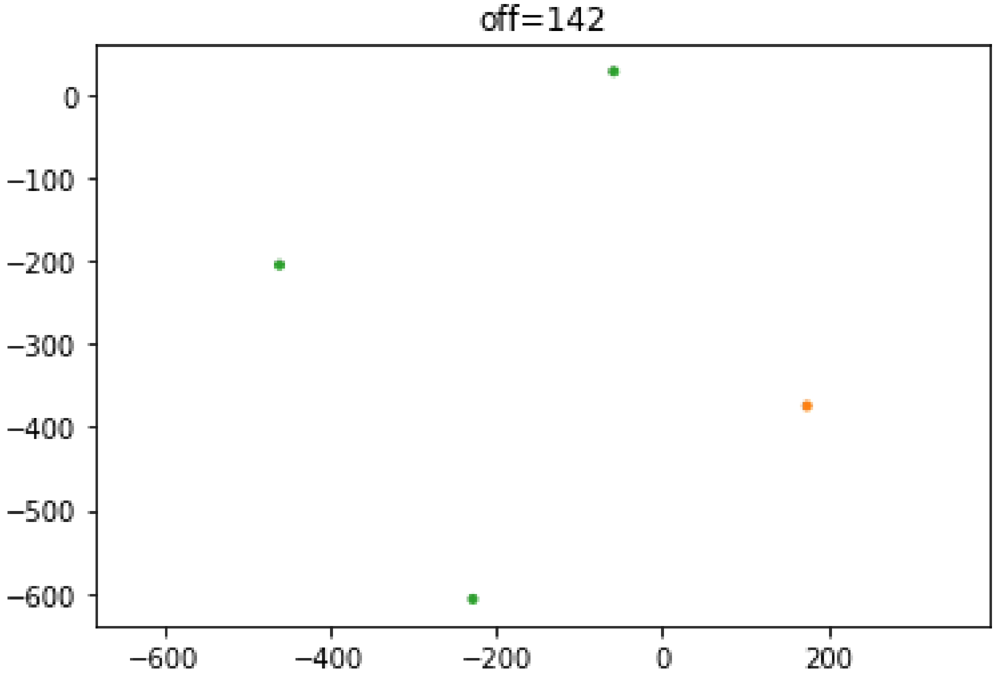

off=135
750
500
250
0
−250
−500
−750
−1000
−500
0
500
1000

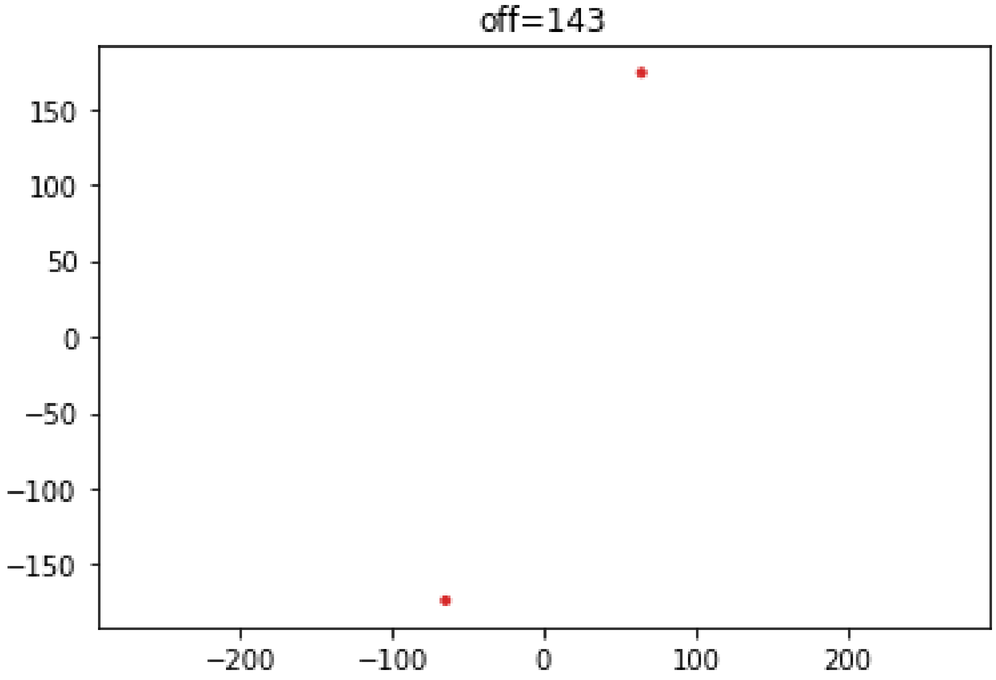

off=136
−300
−350
−400
−450
−400
−350
−300
−250
−200
−150
−100

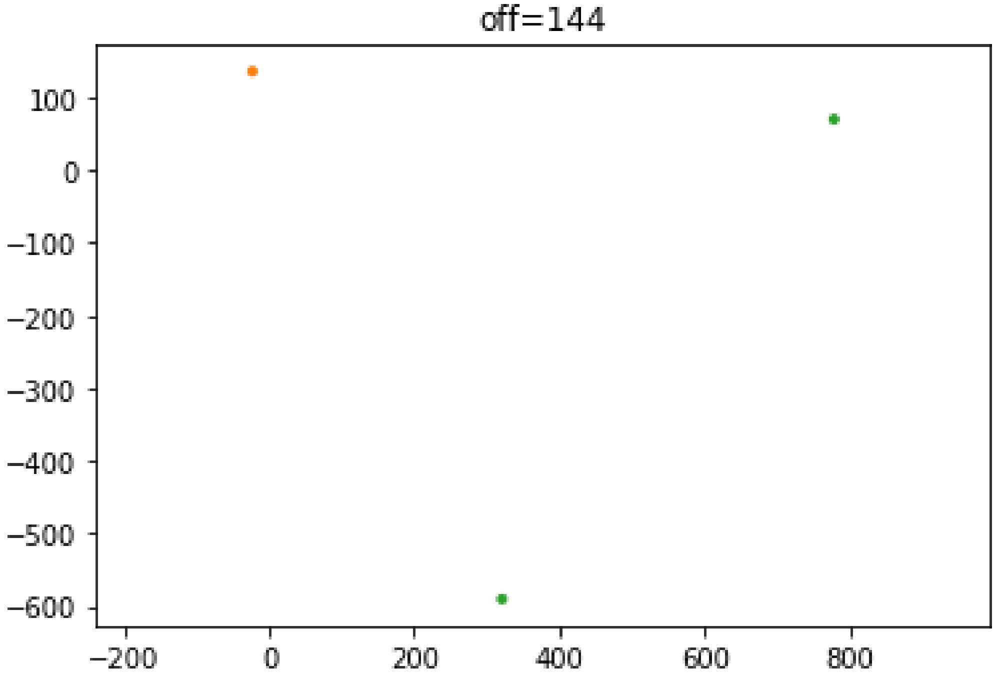

off=137
200
150
100
50
0
150
200
250
300
350
400
450
500

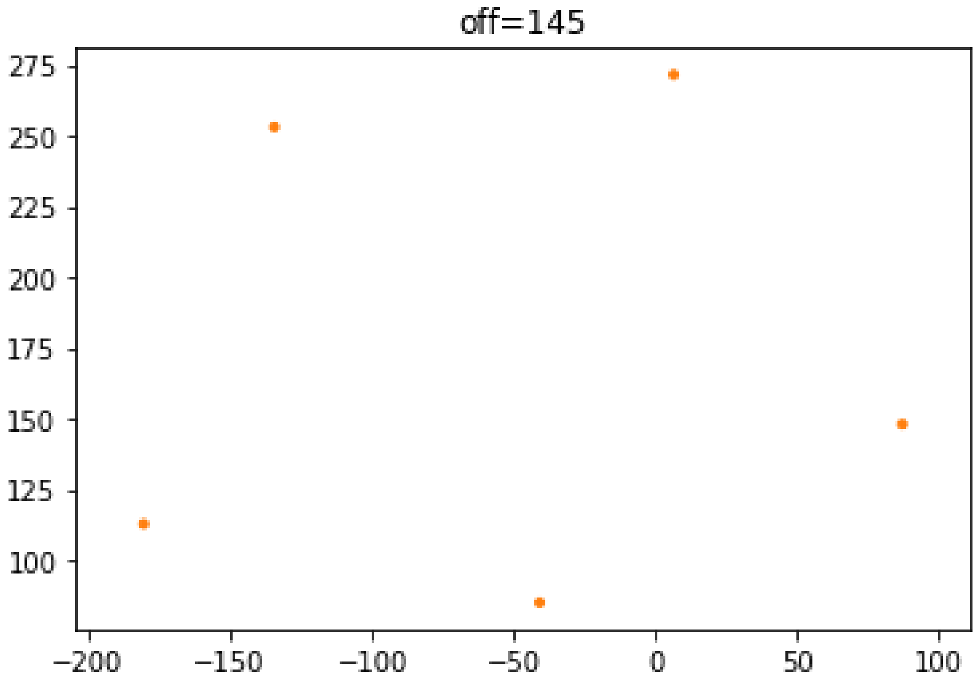

off=139
300
200
100
0
−100
−200
−400
−300
−200
−100
0
100
200
300
400

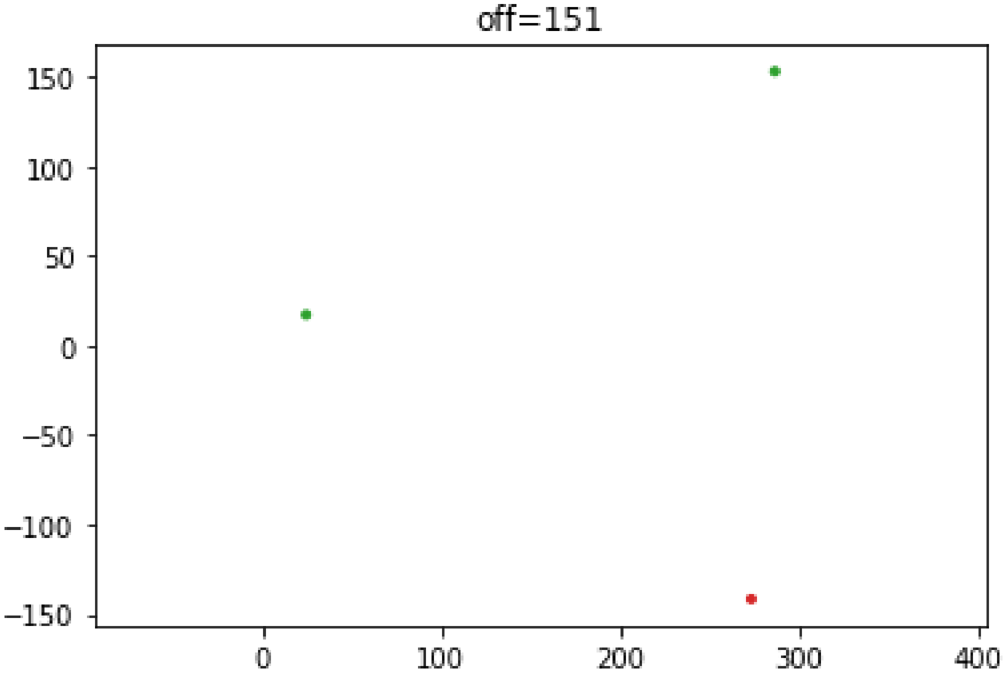

off=140
200
150
100
50
0
−250
−200
−150
−100
−50
0
50
100

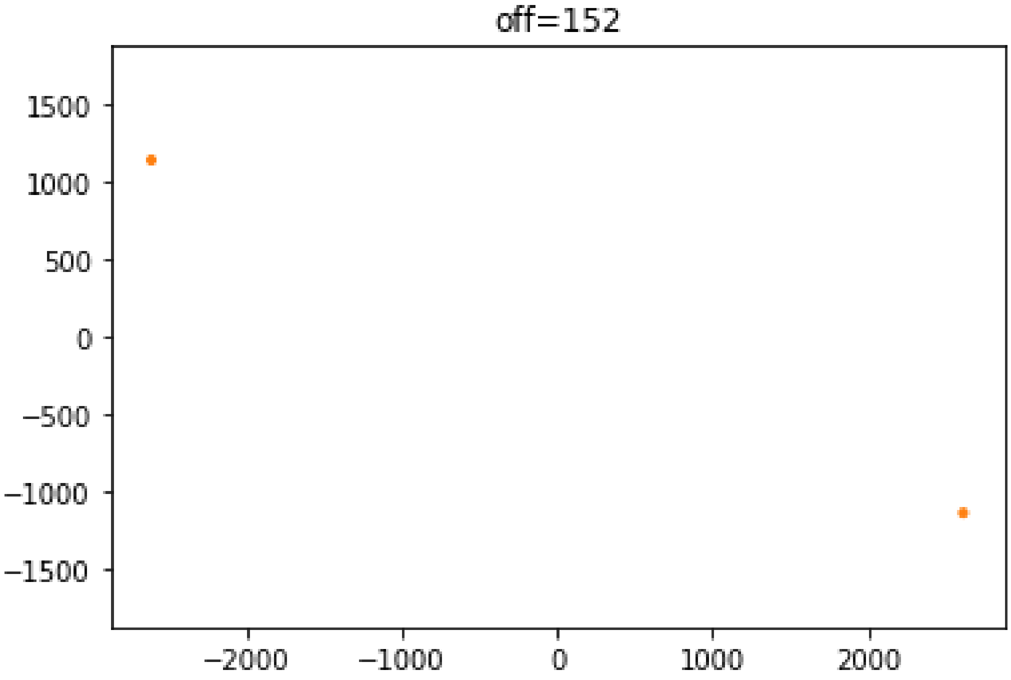

off=142
0
−100
−200
−300
−400
−500
−600
−600
−400
−200
0
200

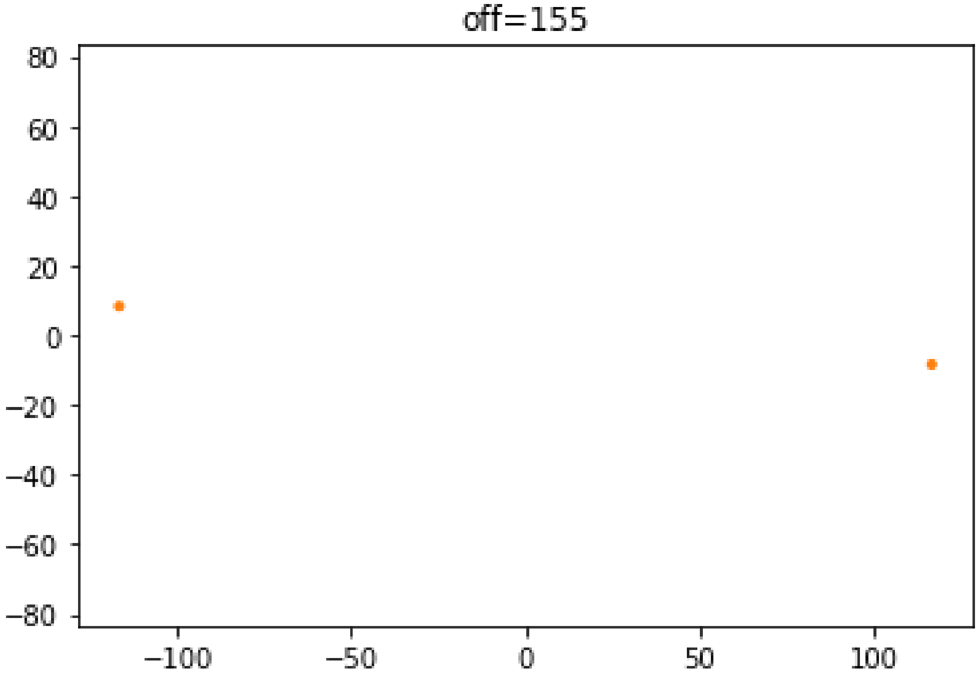

off=143
150
100
50
0
−50
−100
−150
−200
−100
0
100
200

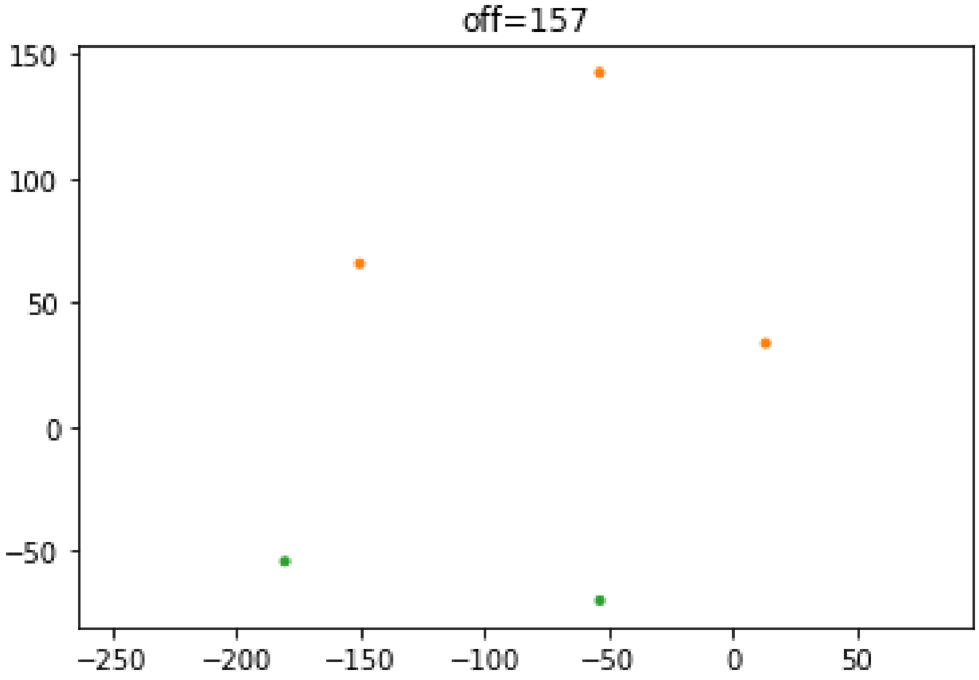

off=144
100
0
−100
−200
−300
−400
−500
−600
−200
0
200
400
600
800

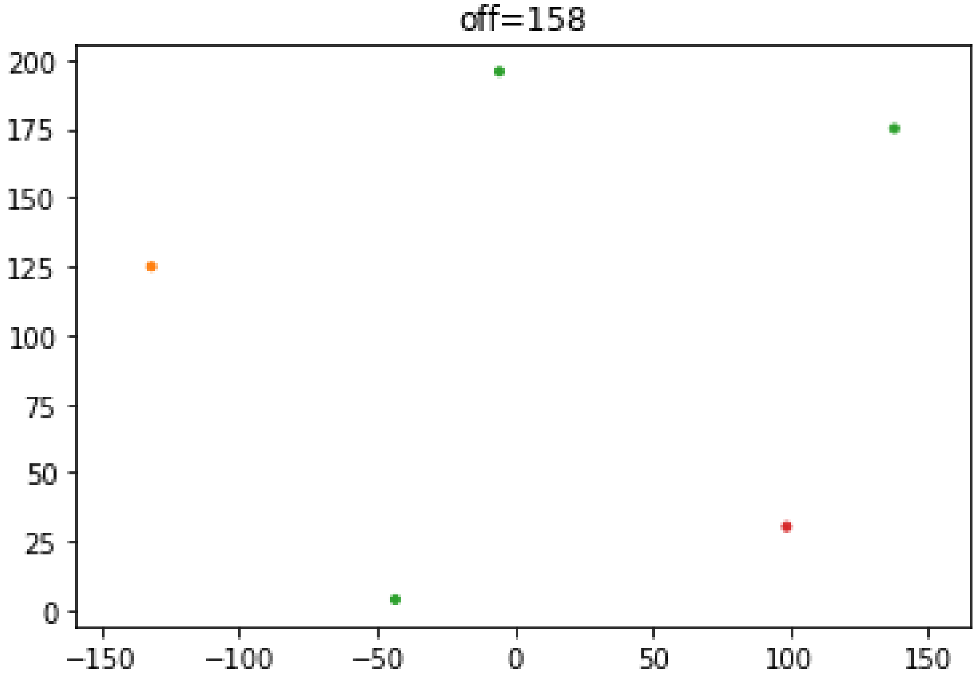

off=145
275
250
225
200
175
150
125
100
−200
−150
−100
−50
0
50
100


off=151
150
100
50
0
−50
−100
−150
0
100
200
300
400


off=152
1500
1000
500
0
−500
−1000
−1500
−2000
−1000
0
1000
2000


off=155
80
60
40
20
0
−20
−40
−60
−80
−100
−50
0
50
100





TSNE is having mized results while UMAP doesn't seem to capture any good groupings. Maybe I should go straight to kmeans or kdtrees\ I should block by crime and examine a few "big" crimes the try several different perplexities.

```
In [ ]:
```