Supplementary materials for Hybrid Spectral/Subspace Clustering of Molecular Dynamics Data

All materials bellow will refer to https://github.com/fio2003/PYSSC/blob/master/additional_data_analysis/segmented_analysis.ods file with analysis.

1 EXTENEDED METHODS

1.1 Spectral Clustering

1.1.1 Concept. Spectral clustering is an algorithm commonly used to cluster data points generated by non-linear processes. In this approach, X is represented as a similarity graph G and is partitioned in such a way that points within one group share high weight (connectivity), while points in different groups share very low weight. To achieve the goal mentioned above, we let graph $G = (V, E, W)$ be an undirected graph with a set of vertices $V = \{v_1, v_2, \ldots, v_N\}$, set of edges $E = \{e_1, e_2, \ldots, e_{N^2}\}$, and weights $W = \{w_{ij}\}_{N\times N}$, $w \geq 0$ shared by every two vertices. A value of $w_{ij} = 0$ will mean that there is no connection between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Each cluster group mentioned above can be described as sets of points A in V that preserve these properties: $A_i \cup A_j = \emptyset$, where $i \neq j$ and $A_1 \cup A_2 \cup \ldots A_n = V$.

In order to clarify the property of similarity within the spectral approach, a brief review of several of the most popular methods for constructing similarity graphs follows:

(1) *Epsilon neighborhood* - points are treated as similar only if their pairwise distance is smaller than a cutoff parameter, epsilon, thus creating an unweighted graph.
(2) *k-nearest neighbors graph* - for each vertex, v_i, k vertices, v_j, are picked with the highest similarity (weight) only if v_i is also among the k-nearest neighbors of v_j, resulting in a mutual k-nearest neighbors graph.
(3) *Fully connected graph* - all vertices in V are connected with a similarity function that encodes all connections.

The Gaussian similarity function (GSF) is one of the most commonly used functions for further refinement of the neighborhood graph and is defined as $s(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/2\sigma^2)$, where σ is a user-defined parameter which determines the rate of decrease in similarity for all points. Selecting an appropriate value for σ can be computationally expensive and time-consuming. Later we will introduce a more advanced method for choosing σ, entropic affinities, that overcomes the restriction of picking a fixed σ in the GSF.

1.1.2 Construction of Graph Laplacians. Once the GSF has been applied to the neighborhood graph, the graph is transformed into Laplacian form. There exist several ways to define graph Laplacians. One of the most common forms used for clustering is the normalized symmetric graph [5], which is defined as:

$$L = D^{-1/2} \times W^* \times D^{-1/2}, \tag{1}$$

where W^* is the similarity matrix formed from the neighborhood graph, with elements defined as $w_{ij} = s(\mathbf{x_i}, \mathbf{x_j})$ and D is a diagonal matrix of size N, where $d_i = \sum_j w_{ij}$. Properties and proofs concerning this Laplacian can be found in [5].

1.1.3 Approximate Normalized Cut. The next step is an application of singular value decomposition (SVD) to the graph Laplacian in order to perform an approximate k-way normalized cut. We save k rows from $V' \in \mathbb{R}^{n\times k}$ (unitary matrix that contains right singular vectors as rows) that correspond to the top k eigenvectors and then construct matrix $Y' \in \mathbb{R}^{n\times k}$ from V' by normalizing the row sums to have norm 1. The matrix Y' represents a nonlinear projection of the data where the clustering problem may be solved using linear clustering algorithms like k-means. The complete procedure for spectral clustering is shown in Algorithm 1.

Algorithm 1 Spectral clustering algorithm

procedure SPECTRAL CLUSTERING(S, k) ▷ Similarity matrix $S \in \mathbb{R}^{n\times n}$, number k of clusters to construct
  Construct a similarity graph by one of the ways described above.
  Compute the normalized Laplacians L using equation 1.
  Compute the first k eigenvectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$ of L.
  Let $V' \in \mathbb{R}^{n\times k}$ be the matrix containing the vectors $v_1, \ldots, v_k$ as columns. Construct matrix $Y' \in \mathbb{R}^{n\times k}$ from V' by normalizing the row sums to have norm 1, that is $y_{ij} = v_{ij}/(\sum_k v_{ik}^2)^{1/2}$.
  Cluster the points ($\mathbf{u_i}$) into clusters $F_1, \ldots, F_k$ with k-means algorithm.
  return Clusters $A_1, \ldots, A_k$ with $A_j = \{j|\boldsymbol{y}_i \in F_j\}$.
end procedure

1.2 Data Subspaces

Clustering of the points in X may be challenging since they may be positioned among a set of m (affine) subspaces $\{S_\ell\}_{\ell=1}^m$ in d dimensions.

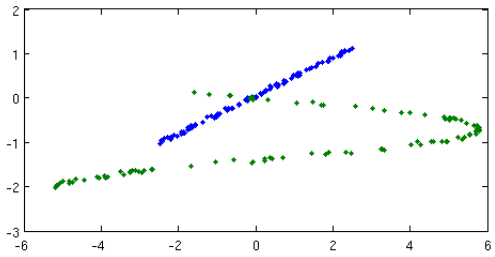

Figure 1: Example of data that resides in two subspaces

Figure 1 demonstrates an intersection of two sets of points that reside in different subspaces, often assumed to have been generated by two distinct latent processes. Clustering data from different subspaces (blue and green) can prove challenging for spectral methods to separate into the desired clusters since the similarity graph will connect points near the region of intersection. Latent processes may also be nonlinear (green) or affine (blue) in nature, but neither the presence of subspaces nor (non)linearity is usually known *a priori* for general data sets. For example, let's say that X resides in d dimensions with m affine subspaces $\{S_\ell\}_{\ell=1}^{m}$. We may often need to cluster X according to the subspaces $\{S_\ell\}_{\ell=1}^{m}$ to obtain good results. In the case of $d = 1$, the problem reduces to the solution of the well-known and easily solved principal components analysis (PCA). However, the problem described above becomes significantly more difficult with the growth of both m and d [2]. Perhaps more importantly, many problems may invoke non-affine (nonlinear) subspaces which may limit the applicability of certain clustering algorithms. In particular, although spectral clustering may solve nonlinear problems, it unfortunately cannot handle clustering within multiple subspaces. For example, in Figure 1 the intersection at (0,0) will not be separated and, most likely, will merge the two data sets in such a way that the normalized cut will not separate the two processes as desired.

1.3 Entropic Affinities

Additionally, picking sigma (σ) for the GSF may be challenging and time consuming. This is especially true when dealing with large data sets of nonuniform density. When applied to data exhibiting both dense and sparse regions, the chosen sigma may keep too many points in dense regions and too few in sparse regions. Entropic affinities promise to overcome such an inconvenience by picking sigma values for each point with respect to a desired perplexity by implicitly defining a continuously differentiable function in the bounded input space [3, 4].

For the posterior distribution of an isotropic kernel density estimator of width σ_i defined on X, let's define a discrete distribution $p_j(\boldsymbol{x_i}; \sigma_i)$ with probabilities for $i, j = 1, \ldots, N$:

$$p_j(x_i; \sigma_i) = \frac{exp\left(-\left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{\sigma_i}\right\|^2\right)}{\sum_{k=1, k!=i}^{N} exp\left(-\left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}_k}{\sigma_i}\right\|^2\right)} \tag{2}$$

In this case each σ_i is being set individually for each point, x_i, to a value such that the entropy of the distribution, $p_j(x_i; \sigma_i)$, equals $log(K)$, where K is a user-set perplexity parameter. Given both the theoretical and practical advantages offered by entropic affinities, we will compare their effectiveness for MD simulation data clustering with the fixed (average) σ approach later on.

1.4 Subspace Clustering

Another approach for clustering, which assumes the data lie in multiple (affine) subspaces was suggested by [2]. It uses the following data property: any point in a union of subspaces can be represented as a linear combination of several points in the local neighborhood. The coefficients of this combination can be used to construct an affinity matrix due to the local relationships between points residing in the same affine subspace. According to the above, X can be viewed as a self-expressive dictionary in which each point $\boldsymbol{x}_i \in \cup_{\ell=1}^{m} \mathcal{S}_\ell$ can be written as a linear combination of other points

$$\boldsymbol{x}_i = X\boldsymbol{c}_i, \tag{3}$$

where $c_i \triangleq [c_{i1}, c_{i2}, \ldots, c_{iN}]^\top$. In order to remove the trivial solution of describing a point as a linear combination of itself, an additional constraint $c_{ii} = 0$ is added.

In practice, the number of data points in a subspace S_ℓ is often higher than its dimension, which concludes that the representation of $\boldsymbol{x}_i$ in the dictionary X is not unique in general. This assumption leads to the conclusion that each $\boldsymbol{x}_i$, and consequently X, has a nontrivial null-space, giving rise to infinitely many representations of each data point.

The key observation in the proposed algorithm was that among all solutions of (3), there exists a sparse solution, $\boldsymbol{c}_i$, whose nonzero entries correspond to data points from the same subspace as $\boldsymbol{x}_i$. Such a solution would be referred as a subspace-sparse representation [2].

A data point $\boldsymbol{x}_i$ that lies in the d_ℓ dimensional subspace S_ℓ can be described as a linear combination of d_ℓ other points from $\mathcal{S}_\ell$. A sparse representation of a data point gives the opportunity to find points from the same subspace where the number of the non-zero elements relates to the dimension of the underlying subspace. A system of equations similar to (3) may contain an infinite number of solutions, which can be restricted by minimization of an appropriate objective function. An example of such a restriction with the ℓ_q-norm of the solution is shown below:

$$min\|\boldsymbol{c}_i\|_q \;\text{ s.t. }\; \boldsymbol{x}_i = X\boldsymbol{c}_i,\; c_{ii} = 0 \tag{4}$$

Different choices of q give rise to different solutions. Usually, by decreasing the value of q from infinity toward zero, the sparsity of the solution increases [2]. The extreme case of $q = 0$ corresponds to the general NP-hard problem of finding the sparsest representation of the given point and is not being considered since we are interested in the efficient way to find a nontrivial sparse representation of $\boldsymbol{x}_i$ in the dictionary X. Minimization of the tightest convex relaxation of the ℓ_1-norm (q=1) is considered as sufficient, which can be solved efficiently using convex programming tools [1].

Equation (4) can be rewritten in matrix form for all data points $i = 1, \ldots, N$ as :

$$min\|\boldsymbol{C}\|_1 \;\text{ s.t. }\; X = XC, \quad diag(C) = 0, \tag{5}$$

where $C \triangleq [c_1 c_2 \ldots c_N] \in \mathbb{R}^{N \times N}$ is the matrix whose ith column corresponds to the sparse representation of $\boldsymbol{x}_i, \boldsymbol{c}_i$, and $diag(\boldsymbol{C}) \in \mathbb{R}^N$ is the vector of the diagonal elements of C. Ideally, the solution of (5) corresponds to sparse subspace representations of the data points, which is used next to derive the clustering of the data.

To perform clustering, first a weighted graph is constructed $\mathcal{G} = (V, E, \boldsymbol{W'})$, where $\boldsymbol{W'} \in \mathbb{R}^{N \times N}$ is a non-negative symmetric similarity matrix representing the weights of the edges. The similarity matrix $\boldsymbol{W'}$, thus the similarity graph $\mathcal{G}$, contains nodes that correspond to the points of the same subspace connected to each other, and there are no edges between nodes that correspond to the points in different subspaces. Let's recall that construction of common graph Laplacians requires a symmetric affinity matrix, while the

sparse representation from the convex optimization does not guarantee symmetry. One possible symmetrization is $\boldsymbol{W'} = |\boldsymbol{C}| + |\boldsymbol{C}|^\top$, which can be described analogously: if node i connected to node j with weight w, then j should have a connection to i with the same weight. The complete procedure for subspace clustering is shown in Algorithm 2. A problematic assumption made using this approach is that the data consists of only *affine* subspaces. We addressed this limitation in the Methods section.

Algorithm 2 Sparse Subspace Clustering

procedure SUBSPACE CLUSTERING(S, k) ▷ A set of points $\{\boldsymbol{x}_i\}_{i=1}^N$ lying in a union of m linear subspaces $\{\mathcal{S}_\ell\}_{\ell=1}^m$.
  Solve the sparse optimization program.
  Normalize the columns of $\boldsymbol{C}$ as $c_i \leftarrow \frac{c_i}{\|c_i\|_\infty}$.
  Compute the first k eigenvectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$ of L.
  Form a similarity graph with N nodes representing the data points. Set the weights on the edges between the nodes by $\boldsymbol{W'} = |\boldsymbol{C}| + |\boldsymbol{C}|^\top$.
  Apply spectral clustering described in Algorithm 1 to the similarity graph $\boldsymbol{W'}$.
  return SpectralClustering($\boldsymbol{W'}, k$). ▷ see Algorithm 1
end procedure

1.4.1 Normalized Mutual Information. Since we need some tool to quantitatively measure each algorithm's ability to separate data into clusters, we utilize normalized mutual information (NMI). Mutual information reflects the dependence of two variables, in our case - simulation replicate number (R) and cluster number (F):

$$NMI(R;F) = \frac{\sum_{f\in F}\sum_{r\in R} p(r,c)log_2\left(\frac{p(r,f)}{p_1(r)p_2(f)}\right)}{argmax(\sum(p(r)log_2p(r);\sum(p(c)log_2p(f))},$$

where $p(r,c)$ is the joint probability distribution of the two random variables R and F; $p_1(r)$ and $p_2(f)$ are the marginal probability distributions of R and F respectively.

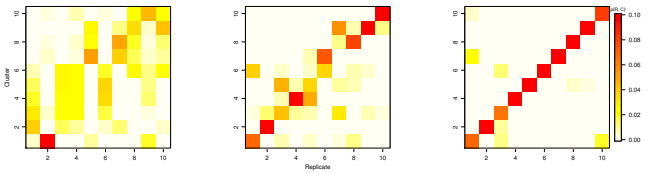

Figure 2: Three examples of cluster-replicate joint probability distributions for low (0.2124, left), medium (0.4495, center), and high (0.6360, right) NMI values.

Examples of different NMI values may be found in Figure 2.

These plots were built with data that can be found in the Results section, but are used here to illustrate the utility of NMI for determining clustering quality for MD simulations. In Figure 2 the left subplot demonstrates the joint probability distribution from Table 2, row 2, column 5, and the center subplot refers to Table 1, row 1, column 1, and the right subplot refers to Table 2, row 7, column 3. For the *high* NMI example (right subplot) we may easily say which data set (replicate simulation) is referred to by a particular cluster, for the medium NMI example (center) we clearly see the trend but it is not always possible to make an assumption about cluster-replicate relationships like we did for example with high NMI value. Finally, the example of low NMI (left subplot) demonstrate an example where the simulation data are too similar so that the algorithm cannot distinguish between the simulations. The critical case of an NMI of 1.0 would exhibit a plot with a filled diagonal that reflects an exact mapping between the input data sets: one per cluster. However, for NMI close to 0.0 we would see an even distribution among the clusters, meaning that the particular algorithm sees no difference between the input data sets. NMI therefore is a good summary of the overall effectiveness of sampling for protein ensembles generated using MD.

2 TEMPERATURE PROFILE

(1) $0 - 20ns$: 300K to 600K
(2) $20 - 80ns$: 600K to 600K
(3) $80 - 100ns$: 600K to 300K
(4) $100 - 350ns$: 300K to 300K

3 THICKNESS EXAMPLE

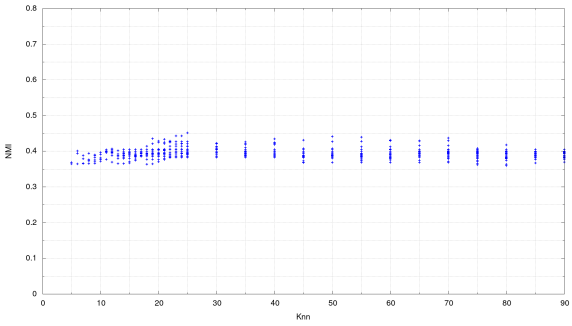

Figure 3: Example of a medium thickness, straight graph derived from NMI/knn results for the SC algorithm with entropic affinities for super sparse data of the TRP Cage protein.

3.1 Boxplot Analysis

3.1.1 Dense. It is clear that SC is the winner for this data set, having general performance much higher than SDS and SES (see Figure 5). We may see that variation among NFPs and NSP1 is about the same, while NUP116 has almost twice the variation. It is also clear that the NFPs had much lower NMI results than IDPs, especially NUP116, as shown in Figure 6.

3.1.2 Sparse. For entropic affinities all three algorithms demonstrate similar variance, but SES generally had higher NMI values. For plain affinities SC had the smallest variance but also the smallest NMI values, while SDS had the highest NMI values, but also the highest variance. For plain affinities SC showed the worst NMI values, but the smallest variation. SDS and SES showed similar variation, but SDS had much higher NMI values (see Figure 7). For

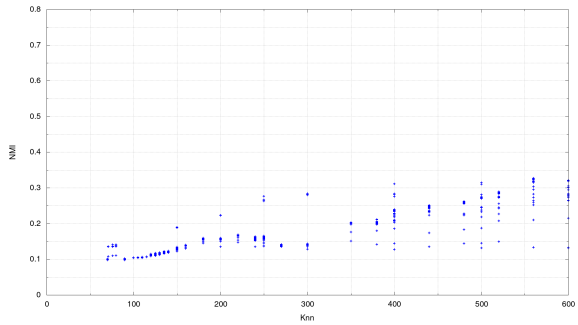

Figure 4: Example of a thickness that changes from medium to wide and has a growing trend; derived from NMI/knn results for the SES algorithm with plain affinity for sparse data of the NSP1 protein.

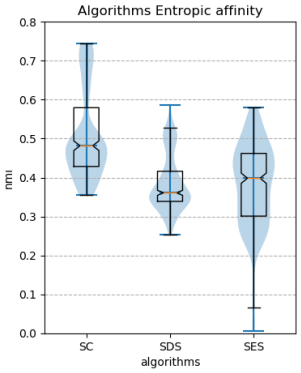

Figure 5: Relationship between NMI values and variation for the SC (left), SDS (middle), and SES (right) algorithms for the perplexity/sigma batch and dense data set.

sparse data and entropic affinities all proteins demonstrated small variance. NFPs had similar NMI values, while NSP1 had a slightly higher NMI value but the highest variation among all proteins and finally NUP116 had the highest NMI values while keeping average variance. For plain affinities all proteins had very high variance. It is interesting that TRP Cage had a much higher (median/average) NMI value than others (see Figure 8).

3.1.3 Super-sparse. For entropic affinities all three algorithms had similar performance, but SC had slightly lower NMI values and SES had slightly higher NMI values. For plain affinities while SC had the smallest variation it also had the smallest NMI values. SDS had medium variation but significantly higher NMI values (see Figure 9). For entropic affinities the NFP group demonstrated very similar results - low variance and lower NMI values. NSP1 had a higher NMI value, but twice the variance, while NUP116 had the highest NMI values and variance. For plain affinities all proteins demonstrated very high variance, but TRP Cage had a slightly higher NMI value (see Figure 10).

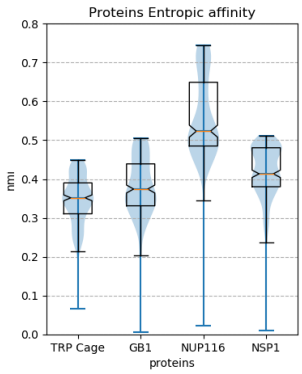

Figure 6: Relationship between NMI values and variation for IDPs: NSP1, NUP116 (left) and NFPs: GB1, TRP Cage (right) for the dense data set.

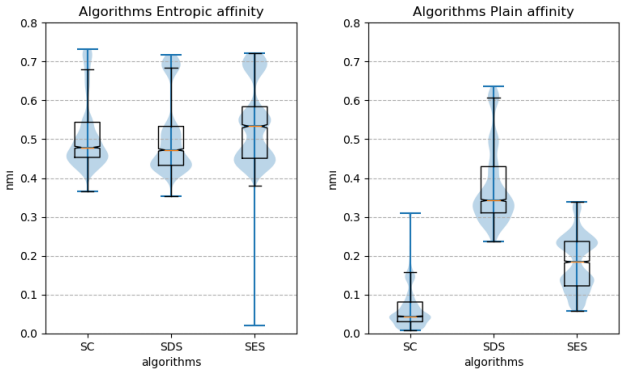

Figure 7: Relationship between NMI values and variation for the SC (left), SDS (middle), and SES (right) algorithms for the perplexity/sigma batch and the sparse data set.

4 PROTEIN TYPE

4.1 Affinities

4.1.1 Width.

EN For knn, NFP showed 65% [AC23] of narrow shapes, 19% [AE23] of wide shapes, and 22% [AF23] of changes. IDP showed 44% [AG23] of narrow shapes, 35% [AI23] of wide shapes, and 56% [AJ23] of changes.
For pp, NFP showed 70% [AC24] of narrow shapes, 6% [AE24] of wide shapes, and 44% [AF24] of changes. IDP showed 65% [AG24] of narrow shapes, 9% [AI24] of wide shapes, and 67% [AJ24] of changes.

PL For knn, NFP showed 36% [AC25] of narrow shapes, 47% [AE25] of wide shapes, and 50% [AF25] of changes. IDP showed 22% [AG25] of narrow shapes, 67% [AI25] of wide shapes, and 83% [AJ25] of changes.
For sigma, NFP showed 14% [AC26] of narrow shapes, 67% [AE26] of wide shapes, and 8% [AF26] of changes. IDP showed

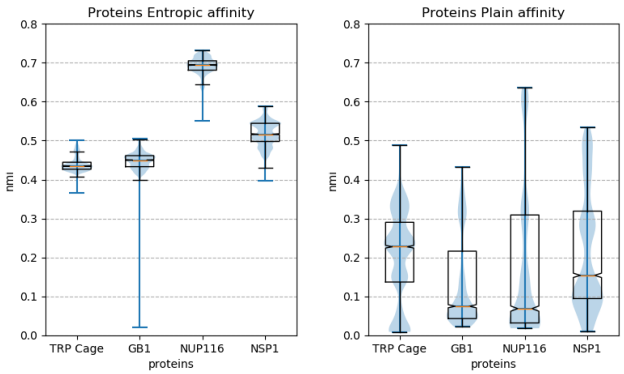

Figure 8: Relationship between NMI values and variation for IDPs: NSP1, NUP116 (left) and NFPs: GB1, TRP Cage (right) for the sparse data set.

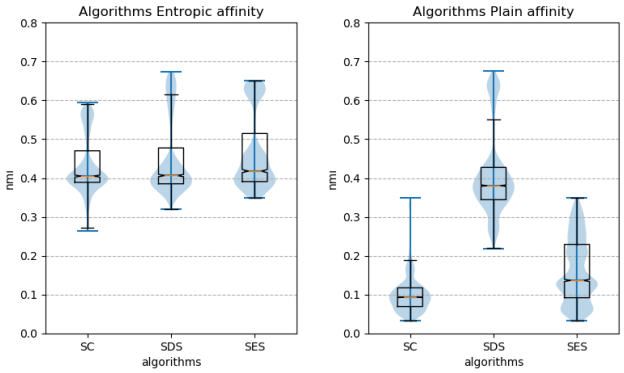

Figure 9: Relationship between NMI values and variation for the SC (left), SDS (middle), and SES (right) algorithms for the perplexity/sigma batch and the super-sparse data set.

0% [AG26] of narrow shapes, 89% [AI26] of wide shapes, and 25% [AJ26] of changes.

Total For knn, NFP showed 53% [AC27] of narrow shapes, 30% [AE27] of wide shapes, and 33% [AF27] of changes. IDP showed 36% [AG27] of narrow shapes, 48% [AI27] of wide shapes, and 67% [AJ27] of changes.
For pp/sigma, NFP showed 48% [AC28] of narrow shapes, 30% [AE28] of wide shapes, and 30% [AF28] of changes. IDP showed 39% [AG28] of narrow shapes, 41% [AI28] of wide shapes, and 50% [AJ28] of changes.

4.1.2 *Shape.*

EN For knn, NFP showed 13% [AC163] of rising parts, 6% [AE163] of falling parts, 6% [AG163]of them were strong, and 17% [AF163] of V and A shapes. IDP showed 17% [AC163] of rising parts, 11% [AE163] of falling parts, 17% [AG163]of them were strong, and 28% [AF163] of V and A shapes.
For pp, NFP showed 9% [AC164] of rising parts, 11% [AE164]

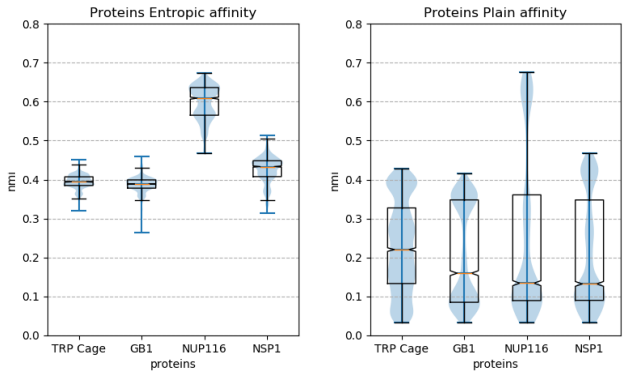

Figure 10: Relationship between NMI values and variation for IDPs: NSP1, NUP116 (left) and NFPs: GB1, TRP Cage (right) for the super-sparse data set.

of falling parts, 0% [AG164]of them were strong, and 22% [AF164] of V and A shapes. IDP showed 13% [AC164] of rising parts, 26% [AE164] of falling parts, 17% [AG164]of them were strong, and 22% [AF164] of V and A shapes.

PL For knn, NFP showed 31% [AC163] of rising parts, 17% [AE163] of falling parts, 58% [AG163]of them were strong, and 50% [AF163] of V and A shapes. IDP showed 33% [AC163] of rising parts, 25% [AE163] of falling parts, 50% [AG163]of them were strong, and 83% [AF163] of V and A shapes.
For sigma, NFP showed 0% [AC164] of rising parts, 6% [AE164] of falling parts, 8% [AG164]of them were strong, and 0% [AF164] of V and A shapes. IDP showed 0% [AC164] of rising parts, 11% [AE164] of falling parts, 33% [AG164]of them were strong, and 33% [AF164] of V and A shapes.

Total For knn, NFP showed 31% [AC163] of rising parts, 17% [AE163] of falling parts, 58% [AG163]of them were strong, and 50% [AF163] of V and A shapes. IDP showed 33% [AC163] of rising parts, 25% [AE163] of falling parts, 50% [AG163]of them were strong, and 83% [AF163] of V and A shapes.
For pp/sigma, NFP showed 0% [AC164] of rising parts, 6% [AE164] of falling parts, 8% [AG164]of them were strong, and 0% [AF164] of V and A shapes. IDP showed 0% [AC164] of rising parts, 11% [AE164] of falling parts, 33% [AG164]of them were strong, and 33% [AF164] of V and A shapes.

4.2 Algorithms

4.2.1 *Width.*

SC For knn, NFP showed 27% [AK87] of narrow shapes, 47% [AM87] of wide shapes, and 30% [AF91] of changes. IDP showed 20% [AK93] of narrow shapes, 57% [AM93] of wide shapes, and 90% [AN93] of changes.
For pp/sigma, NFP showed 60% [AK88] of narrow shapes, 30% [AM88] of wide shapes, and 20% [AN88] of changes. IDP showed 43% [AK94] of narrow shapes, 27% [AM94] of wide shapes, and 70% [AN94] of changes.

SDS For knn, NFP showed 60% [AK89] of narrow shapes, 20% [AK89] of wide shapes, and 30% [AN89] of changes. IDP showed 43% [AC95] of narrow shapes, 47% [AM95] of wide shapes, and 40% [AN95] of changes.
For pp/sigma, NFP showed 40% [AK90] of narrow shapes, 20% [AM90] of wide shapes, and 40% [AN90] of changes. IDP showed 40% [AK96] of narrow shapes, 47% [AM96] of wide shapes, and 50% [AN96] of changes.

SES For knn, NFP showed 73% [AK91] of narrow shapes, 23% [AM91] of wide shapes, and 40% [AN91] of changes. IDP showed 43% [AK97] of narrow shapes, 40% [AM97] of wide shapes, and 70% [AN97] of changes.
For pp/sigma, NFP showed 43% [AK92] of narrow shapes, 40% [AM92] of wide shapes, and 30% [AN92] of changes. IDP showed 33% [AK98] of narrow shapes, 50% [AM98] of wide shapes, and 30% [AN98] of changes.

Total For knn, NFP showed 53% [AK105] of narrow shapes, 30% [AM105] of wide shapes, and 33% [AN105] of changes. IDP showed 36% [AK107] of narrow shapes, 48% [AM107] of wide shapes, and 67% [AN107] of changes.
For pp/sigma, NFP showed 48% [AK106] of narrow shapes, 30% [AM106] of wide shapes, and 30% [AN106] of changes. IDP showed 39% [AK108] of narrow shapes, 41% [AM108] of wide shapes, and 50% [AN108] of changes.

4.2.2 *Shape.*

SC For knn, NFP showed 10% [AC241] of rising parts, 3% [AE241] of falling parts, 20% [AG241]of them were strong, and 30% [AF241] of V and A shapes. IDP showed 7% [AC243] of rising parts, 13% [AE243] of falling parts, 0% [AG243]of them were strong, and 50% [AF243] of V and A shapes.
For pp/sigma, NFP showed 7% [AC242] of rising parts, 7% [AE242] of falling parts, 0% [AG242]of them were strong, and 10% [AF242] of V and A shapes. IDP showed 7% [AC244] of rising parts, 27% [AE244] of falling parts, 20% [AG244]of them were strong, and 50% [AF244] of V and A shapes.

SDS For knn, NFP showed 7% [AH241] of rising parts, 20% [AJ241] of falling parts, 20% [AL241]of them were strong, and 40% [AK241] of V and A shapes. IDP showed 7% [AH243] of rising parts, 30% [AJ243] of falling parts, 40% [AL243]of them were strong, and 60% [AK243] of V and A shapes.
For pp/sigma, NFP showed 3% [AH242] of rising parts, 10% [AJ242] of falling parts, 10% [AL242]of them were strong, and 10% [AK242] of V and A shapes. IDP showed 7% [AH244] of rising parts, 20% [AJ244] of falling parts, 40% [AL244]of them were strong, and 30% [AK244] of V and A shapes.

SES For knn, NFP showed 43% [AM241] of rising parts, 7% [AO241] of falling parts, 40% [AQ241]of them were strong, and 20% [AP241] of V and A shapes. IDP showed 57% [AM243] of rising parts, 7% [AO243] of falling parts, 50% [AQ243]of them were strong, and 40% [AP243] of V and A shapes.
For pp/sigma, NFP showed 7% [AM242] of rising parts, 10% [AO242] of falling parts, 0% [AQ242]of them were strong, and 20% [AP242] of V and A shapes. IDP showed 10% [AM244] of rising parts, 13% [AO244] of falling parts, 10% [AQ244]of them were strong, and 0% [AP244] of V and A shapes.

Total For knn, NFP showed 20% [AR241] of rising parts, 10% [AT241] of falling parts, 27% [AV241]of them were strong, and 30% [AU241] of V and A shapes. IDP showed 57% [AR243] of rising parts, 7% [AT243] of falling parts, 50% [AV243]of them were strong, and 40% [AU243] of V and A shapes.
For pp/sigma, NFP showed 6% [AR242] of rising parts, 9% [AT242] of falling parts, 3% [AV242]of them were strong, and 13% [AU242] of V and A shapes. IDP showed 8% [AR244] of rising parts, 20% [AT244] of falling parts, 23% [AV244]of them were strong, and 27% [AU244] of V and A shapes.

4.3 Sparsity

4.3.1 *Width.*

DN For knn, NFP showed 33% [AO59] of narrow shapes, 39% [AQ59] of wide shapes, and 50% [AR59] of changes. IDP showed 33% [AO65] of narrow shapes, 44% [AQ65] of wide shapes, and 83% [AR65] of changes.
For pp, NFP showed 56% [AO60] of narrow shapes, 17% [AQ60] of wide shapes, and 50% [AR60] of changes. IDP showed 50% [AO66] of narrow shapes, 22% [AQ66] of wide shapes, and 83% [AR66] of changes.

SP For knn, NFP showed 53% [AO61] of narrow shapes, 36% [AQ61] of wide shapes, and 33% [AR61] of changes. IDP showed 33% [AO67] of narrow shapes, 50% [AQ67] of wide shapes, and 50% [AR67] of changes.
For pp/sigma, NFP showed 50% [AO62] of narrow shapes, 25% [AQ62] of wide shapes, and 33% [AR62] of changes. IDP showed 36% [AO68] of narrow shapes, 39% [AQ68] of wide shapes, and 42% [AR68] of changes.

SS For knn, NFP showed 64% [AO63] of narrow shapes, 19% [AQ63] of wide shapes, and 25% [AR63] of changes. IDP showed 39% [AO69] of narrow shapes, 47% [AQ69] of wide shapes, and 75% [AR69] of changes.
For pp/sigma, NFP showed 42% [AO64] of narrow shapes, 42% [AQ64] of wide shapes, and 17% [AR64] of changes. IDP showed 36% [AO70] of narrow shapes, 53% [AQ70] of wide shapes, and 42% [AR70] of changes.

Total For knn, NFP showed 53% [AO77] of narrow shapes, 30% [AQ77] of wide shapes, and 33% [AR77] of changes. IDP showed 36% [AO79] of narrow shapes, 48% [AQ79] of wide shapes, and 67% [AR79] of changes.
For pp/sigma, NFP showed 48% [AO78] of narrow shapes, 30% [AQ78] of wide shapes, and 30% [AR78] of changes. IDP showed 39% [AO80] of narrow shapes, 41% [AQ80] of wide shapes, and 50% [AR80] of changes.

4.3.2 *Shape.*

DN For knn, NFP showed 28% [AR223] of rising parts, 17% [AT223] of falling parts, 17% [AV223]of them were strong, and 50% [AU223] of V and A shapes. IDP showed 22% [AR229] of rising parts, 11% [AT229] of falling parts, 33% [AV229]of them were strong, and 50% [AU229] of V and A shapes.
For pp, NFP showed 11% [AR224] of rising parts, 17% [AT224] of falling parts, 0% [AV224]of them were strong, and 17%

[AU224] of V and A shapes. IDP showed 11% [AR230] of rising parts, 50% [AT230] of falling parts, 33% [AV230]of them were strong, and 17% [AU230] of V and A shapes.

SP For knn, NFP showed 17% [AR225] of rising parts, 8% [AT225] of falling parts, 17% [AV225]of them were strong, and 25% [AU225] of V and A shapes. IDP showed 22% [AR231] of rising parts, 19% [AT231] of falling parts, 17% [AV231]of them were strong, and 58% [AU231] of V and A shapes.
For pp/sigma, NFP showed 6% [AR226] of rising parts, 11% [AT226] of falling parts, 17% [AV226]of them were strong, and 17% [AU226] of V and A shapes. IDP showed 3% [AR232] of rising parts, 14% [AT232] of falling parts, 25% [AV232]of them were strong, and 25% [AU232] of V and A shapes.

SS For knn, NFP showed 19% [AR227] of rising parts, 8% [AT227] of falling parts, 42% [AV227]of them were strong, and 25% [AU227] of V and A shapes. IDP showed 25% [AR233] of rising parts, 17% [AT233] of falling parts, 42% [AV233]of them were strong, and 42% [AU233] of V and A shapes.
For pp/sigma, NFP showed 3% [AR228] of rising parts, 3% [AT228] of falling parts, 0% [AV228]of them were strong, and 8% [AU228] of V and A shapes. IDP showed 11% [AR234] of rising parts, 11% [AT234] of falling parts, 17% [AV234]of them were strong, and 33% [AU234] of V and A shapes.

Total For knn, NFP showed 20% [AR241] of rising parts, 10% [AT241] of falling parts, 27% [AV241]of them were strong, and 30% [AU241] of V and A shapes. IDP showed 23% [AR243] of rising parts, 17% [AT243] of falling parts, 30% [AV243]of them were strong, and 50% [AU243] of V and A shapes.
For pp/sigma, NFP showed 6% [AR242] of rising parts, 9% [AT242] of falling parts, 3% [AV242]of them were strong, and 13% [AU242] of V and A shapes. IDP showed 8% [AR244] of rising parts, 20% [AT244] of falling parts, 23% [AV244]of them were strong, and 27% [AU244] of V and A shapes.

5 AFFINITY TYPE

5.1 Protein type

5.1.1 *Width.*

NFP For knn, Entropic affinities showed 65% [AC23] of narrow shapes, 19% [AE23] of wide shapes, and 22% [AF23] of changes. Plain affinities showed 36% [AC25] of narrow shapes, 47% [AE25] of wide shapes, and 50% [AF25] of changes.
For pp/sigma, Entropic affinities showed 70% [AC24] of narrow shapes, 6% [AE24] of wide shapes, and 44% [AF24] of changes. Plain affinities showed 14% [AC26] of narrow shapes, 67% [AE26] of wide shapes, and 8% [AF26] of changes.

IDP For knn, Entropic affinities showed 44% [AG23] of narrow shapes, 35% [AI23] of wide shapes, and 56% [AJ23] of changes. Plain affinities showed 22% [AG25] of narrow shapes, 67% [AI25] of wide shapes, and 83% [AJ25] of changes.
For pp/sigma, Entropic affinities showed 65% [AG24] of narrow shapes, 9% [AI24] of wide shapes, and 67% [AJ24] of changes. Plain affinities showed 0% [AG26] of narrow shapes, 89% [AI26] of wide shapes, and 25% [AJ26] of changes.

Total For knn, Entropic affinities showed 55% [AK23] of narrow shapes, 27% [AM23] of wide shapes, and 39% [AN23] of changes. Plain affinities showed 29% [AK25] of narrow shapes, 57% [AM25] of wide shapes, and 67% [AN25] of changes.
For pp/sigma, Entropic affinities showed 68% [AK24] of narrow shapes, 7% [AM24] of wide shapes, and 56% [AN24] of changes. Plain affinities showed 7% [AK26] of narrow shapes, 78% [AM26] of wide shapes, and 17% [AN26] of changes.

5.1.2 *Shape.*

NFP For knn, Entropic affinities showed 13% [AC267] of rising parts, 6% [AE267] of falling parts, 6% [AG267]of them were strong, and 17% [AF267] of V and A shapes. Plain affinities showed 31% [AH267] of rising parts, 17% [AJ267] of falling parts, 58% [AL267]of them were strong, and 50% [AK267] of V and A shapes.
For pp/sigma, Entropic affinities showed 9% [AC268] of rising parts, 11% [AE268] of falling parts, 0% [AG268]of them were strong, and 22% [AF268] of V and A shapes. Plain affinities showed 0% [AH268] of rising parts, 6% [AJ268] of falling parts, 8% [AL268]of them were strong, and 0% [AK268] of V and A shapes.

IDP For knn, Entropic affinities showed 17% [AC269] of rising parts, 11% [AE269] of falling parts, 17% [AG269]of them were strong, and 28% [AF269] of V and A shapes. Plain affinities showed 33% [AH269] of rising parts, 25% [AJ269] of falling parts, 50% [AL269]of them were strong, and 83% [AK269] of V and A shapes.
For pp/sigma, Entropic affinities showed 13% [AC270] of rising parts, 26% [AE270] of falling parts, 17% [AG270]of them were strong, and 22% [AF270] of V and A shapes. Plain affinities showed 0% [AH270] of rising parts, 11% [AJ270] of falling parts, 33% [AL270]of them were strong, and 33% [AK270] of V and A shapes.

Total For knn, Entropic affinities showed 15% [AC271] of rising parts, 8% [AE271] of falling parts, 11% [AG271]of them were strong, and 22% [AF271] of V and A shapes. Plain affinities showed 32% [AH271] of rising parts, 21% [AJ271] of falling parts, 54% [AL271]of them were strong, and 67% [AK271] of V and A shapes.
For pp/sigma, Entropic affinities showed 11% [AC272] of rising parts, 19% [AE272] of falling parts, 8% [AG272]of them were strong, and 22% [AF272] of V and A shapes. Plain affinities showed 0% [AH272] of rising parts, 8% [AJ272] of falling parts, 21% [AL272]of them were strong, and 17% [AK272] of V and A shapes.

5.2 Algorithms

5.2.1 *Width.*

SC For knn, Entropic affinities showed 19% [AC99] of narrow shapes, 47% [AE99] of wide shapes, and 42% [AF99] of changes. Plain affinities showed 29% [AF99] of narrow shapes, 58% [AI99] of wide shapes, and 88% [AJ99] of changes.
For pp/sigma, Entropic affinities showed 81% [AC100] of narrow shapes, 0% [AE100] of wide shapes, and 50% [AF100] of changes. Plain affinities showed 8% [AG100] of narrow shapes, 71% [AI100] of wide shapes, and 38% [AJ100] of changes.

SDS For knn, Entropic affinities showed 81% [AC101] of narrow shapes, 8% [AE101] of wide shapes, and 25% [AF101] of changes. Plain affinities showed 8% [AF101] of narrow shapes, 71% [AI101] of wide shapes, and 50% [AJ101] of changes.
For pp/sigma, Entropic affinities showed 67% [AC102] of narrow shapes, 6% [AE101] of wide shapes, and 67% [AF102] of changes. Plain affinities showed 0% [AG102] of narrow shapes, 75% [AI102] of wide shapes, and 13% [AJ102] of changes.

SES For knn, Entropic affinities showed 64% [AC103] of narrow shapes, 25% [AE103] of wide shapes, and 50% [AF103] of changes. Plain affinities showed 50% [AF103] of narrow shapes, 42% [AI103] of wide shapes, and 63% [AJ103] of changes.
For pp/sigma, Entropic affinities showed 56% [AC104] of narrow shapes, 17% [AE104] of wide shapes, and 50% [AF104] of changes. Plain affinities showed 12.5% [AG104] of narrow shapes, 87.5% [AI100] of wide shapes, and 0% [AJ100] of changes.

Total For knn, Entropic affinities showed 55% [AC109] of narrow shapes, 27% [AE109] of wide shapes, and 39% [AF109] of changes. Plain affinities showed 29% [AF109] of narrow shapes, 57% [AI109] of wide shapes, and 67% [AJ109] of changes.
For pp/sigma, Entropic affinities showed 68% [AC109] of narrow shapes, 7% [AE109] of wide shapes, and 56% [AF109] of changes. Plain affinities showed 7% [AG109] of narrow shapes, 78% [AI109] of wide shapes, and 17% [AJ109] of changes.

5.2.2 *Shape.*

SC For knn, Entropic affinities showed 3% [AC261] of rising parts, 6% [AE261] of falling parts, 0% [AG261]of them were strong, and 8% [AF261] of V and A shapes. Plain affinities showed 17% [AH261] of rising parts, 13% [AJ261] of falling parts, 25% [AL261]of them were strong, and 88% [AK261] of V and A shapes.
For pp/sigma, Entropic affinities showed 11% [AC262] of rising parts, 28% [AE262] of falling parts, 17% [AG262]of them were strong, and 33% [AF262] of V and A shapes. Plain affinities showed 0% [AH262] of rising parts, 0% [AJ262] of falling parts, 0% [AL262]of them were strong, and 25% [AK262] of V and A shapes.

SDS For knn, Entropic affinities showed 8% [AC263] of rising parts, 8% [AE263] of falling parts, 8% [AG263]of them were strong, and 25% [AF263] of V and A shapes. Plain affinities showed 4% [AH263] of rising parts, 50% [AJ263] of falling parts, 63% [AL263]of them were strong, and 88% [AK263] of V and A shapes.
For pp/sigma, Entropic affinities showed 8% [AC264] of rising parts, 8% [AE264] of falling parts, 0% [AG264]of them were strong, and 17% [AF264] of V and A shapes. Plain affinities showed 0% [AH264] of rising parts, 25% [AJ264] of falling parts, 63% [AL264]of them were strong, and 25% [AK264] of V and A shapes.

SES For knn, Entropic affinities showed 33% [AC265] of rising parts, 11% [AE265] of falling parts, 25% [AG265]of them were strong, and 33% [AF265] of V and A shapes. Plain affinities showed 75% [AH265] of rising parts, 0% [AJ265] of falling parts, 75% [AL265]of them were strong, and 25% [AK265] of V and A shapes.
For pp/sigma, Entropic affinities showed 14% [AC266] of rising parts, 19% [AE266] of falling parts, 8% [AG266]of them were strong, and 17% [AF266] of V and A shapes. Plain affinities showed 0% [AH266] of rising parts, 0% [AJ266] of falling parts, 0% [AL266]of them were strong, and 0% [AK266] of V and A shapes.

Total For knn, Entropic affinities showed 15% [AC271] of rising parts, 8% [AE271] of falling parts, 11% [AG271]of them were strong, and 22% [AF271] of V and A shapes. Plain affinities showed 32% [AH271] of rising parts, 21% [AJ271] of falling parts, 54% [AL271]of them were strong, and 67% [AK271] of V and A shapes.
For pp/sigma, Entropic affinities showed 11% [AC272] of rising parts, 19% [AE272] of falling parts, 8% [AG272]of them were strong, and 22% [AF272] of V and A shapes. Plain affinities showed 0% [AH272] of rising parts, 8% [AJ272] of falling parts, 21% [AL272]of them were strong, and 17% [AK272] of V and A shapes.

5.3 Data sparsity

5.3.1 *Width.*

DN For knn, Entropic affinities showed 33% [AK5] of narrow shapes, 42% [AM5] of wide shapes, and 67% [AN5] of changes.
For pp, Entropic affinities showed 53% [AK6] of narrow shapes, 19% [AM6] of wide shapes, and 67% [AN6] of changes.

SP For knn, Entropic affinities showed 69% [AK7] of narrow shapes, 8% [AM7] of wide shapes, and 25% [AN7] of changes. Plain affinities showed 17% [AK13] of narrow shapes, 78% [AM13] of wide shapes, and 58% [AN13] of changes.
For pp/sigma, Entropic affinities showed 72% [AK8] of narrow shapes, 0% [AM8] of wide shapes, and 50% [AN8] of changes. Plain affinities showed 14% [AK14] of narrow shapes, 64% [AM14] of wide shapes, and 25% [AN14] of changes.

SS For knn, Entropic affinities showed 61% [AK9] of narrow shapes, 31% [AM9] of wide shapes, and 25% [AN9] of changes. Plain affinities showed 42% [AK15] of narrow shapes, 36% [AM15] of wide shapes, and 75% [AN15] of changes.
For pp/sigma, Entropic affinities showed 78% [AK10] of narrow shapes, 3% [AM10] of wide shapes, and 50% [AN10] of changes. Plain affinities showed 0% [AK16] of narrow shapes, 92% [AM16] of wide shapes, and 8% [AN16] of changes.

Total For knn, Entropic affinities showed 55% [AK23] of narrow shapes, 27% [AM23] of wide shapes, and 39% [AN23] of changes. Plain affinities showed 29% [AK25] of narrow shapes, 57% [AM25] of wide shapes, and 67% [AN25] of changes.
For pp/sigma, Entropic affinities showed 68% [AK24] of narrow shapes, 7% [AM24] of wide shapes, and 56% [AN24] of changes. Plain affinities showed 7% [AK26] of narrow shapes, 78% [AM26] of wide shapes, and 17% [AN26] of changes.

5.3.2 Shape.

DN For knn, Entropic affinities showed 25% [AM145] of rising parts, 14% [AO145] of falling parts, 25% [AQ145]of them were strong, and 50% [AP145] of V and A shapes. For pp, Entropic affinities showed 11% [AM146] of rising parts, 33% [AO146] of falling parts, 17% [AQ146]of them were strong, and 17% [AP146] of V and A shapes.

SP For knn, Entropic affinities showed 14% [AM147] of rising parts, 8% [AO147] of falling parts, 0% [AQ147]of them were strong, and 17% [AP147] of V and A shapes. Plain affinities showed 25% [AM153] of rising parts, 19% [AO153] of falling parts, 33% [AQ153]of them were strong, and 67% [AP153] of V and A shapes.
For pp/sigma, Entropic affinities showed 8% [AM148] of rising parts, 17% [AO148] of falling parts, 8% [AQ148]of them were strong, and 25% [AP148] of V and A shapes. Plain affinities showed 0% [AM154] of rising parts, 8% [AO154] of falling parts, 25% [AQ154]of them were strong, and 17% [AP154] of V and A shapes.

SS For knn, Entropic affinities showed 6% [AM149] of rising parts, 3% [AO149] of falling parts, 8% [AQ149]of them were strong, and 0% [AP149] of V and A shapes. Plain affinities showed 39% [AM155] of rising parts, 22% [AO155] of falling parts, 75% [AQ155]of them were strong, and 67% [AP155] of V and A shapes.
For pp/sigma, Entropic affinities showed 14% [AM150] of rising parts, 6% [AO150] of falling parts, 0% [AQ150]of them were strong, and 25% [AP150] of V and A shapes. Plain affinities showed 0% [AM156] of rising parts, 8% [AO156] of falling parts, 17% [AQ156]of them were strong, and 17% [AP156] of V and A shapes.

Total For knn, Entropic affinities showed 15% [AM149] of rising parts, 8% [AO149] of falling parts, 11% [AQ149]of them were strong, and 22% [AP149] of V and A shapes. Plain affinities showed 32% [AM155] of rising parts, 21% [AO155] of falling parts, 54% [AQ155]of them were strong, and 67% [AP155] of V and A shapes.
For pp/sigma, Entropic affinities showed 11% [AM150] of rising parts, 19% [AO150] of falling parts, 8% [AQ150]of them were strong, and 22% [AP150] of V and A shapes. Plain affinities showed 0% [AM156] of rising parts, 8% [AO156] of falling parts, 21% [AQ156]of them were strong, and 17% [AP156] of V and A shapes.

6 ALGORITHM TYPE

6.1 Protein

6.1.1 Width.

NFP For knn, SC algorithm showed 27% [AK87] of narrow shapes, 47% [AM87] of wide shapes, and 30% [AN87] of changes. SDS algorithm showed 60% [AK89] of narrow shapes, 20% [AM89] of wide shapes, and 30% [AN89] of changes. SES algorithm showed 73% [AK91] of narrow shapes, 23% [AM91] of wide shapes, and 40% [AN91] of changes. For pp/sigma, SC algorithm showed 60% [AK88] of narrow shapes, 30% [AM88] of wide shapes, and 20% [AN88] of changes. SDS algorithm showed 40% [AK90] of narrow shapes, 20% [AM90] of wide shapes, and 40% [AN90] of changes. SES algorithm showed 43% [AK92] of narrow shapes, 40% [AM92] of wide shapes, and 30% [AN92] of changes.

IDP For knn, SC algorithm showed 20% [AK93] of narrow shapes, 57% [AM93] of wide shapes, and 90% [AN93] of changes. SDS algorithm showed 43% [AK95] of narrow shapes, 47% [AM95] of wide shapes, and 40% [AN95] of changes. SES algorithm showed 43% [AK97] of narrow shapes, 40% [AM97] of wide shapes, and 70% [AN97] of changes.
For pp/sigma, SC algorithm showed 43% [AK94] of narrow shapes, 27% [AM94] of wide shapes, and 70% [AN94] of changes. SDS algorithm showed 40% [AK96] of narrow shapes, 47% [AM96] of wide shapes, and 50% [AN96] of changes. SES algorithm showed 33% [AK98] of narrow shapes, 50% [AM98] of wide shapes, and 30% [AN98] of changes.

Total For knn, SC algorithm showed 23% [AK99] of narrow shapes, 52% [AM99] of wide shapes, and 60% [AN99] of changes. SDS algorithm showed 52% [AK101] of narrow shapes, 33% [AM101] of wide shapes, and 35% [AN101] of changes. SES algorithm showed 58% [AK103] of narrow shapes, 32% [AM103] of wide shapes, and 55% [AN103] of changes.
For pp/sigma, SC algorithm showed 52% [AK100] of narrow shapes, 28% [AM100] of wide shapes, and 45% [AN100] of changes. SDS algorithm showed 40% [AK102] of narrow shapes, 33% [AM102] of wide shapes, and 45% [AN102] of changes. SES algorithm showed 38% [AK104] of narrow shapes, 45% [AM104] of wide shapes, and 30% [AN104] of changes.

6.1.2 Shape.

NFP For knn, SC algorithm showed 17% [AH249] of rising parts, 8% [AJ249] of falling parts, 50% [AL249] of them were strong, and 75% [AK249] of V and A shapes. SDS algorithm showed 8% [AH251] of rising parts, 42% [AJ251] of falling parts, 50% [AL251] of them were strong, and 75% [AK251] of V and A shapes. SES algorithm showed 67% [AH253] of rising parts, 0% [AJ253] of falling parts, 75% [AL253] of them were strong, and 0% [AK253] of V and A shapes.
For pp/sigma, SC algorithm showed 0% [AH250] of rising parts, 0% [AJ250] of falling parts, 0% [AL250] of them were strong, and 0% [AK250] of V and A shapes. SDS algorithm showed 0% [AH252] of rising parts, 17% [AJ252] of falling parts, 25% [AL252] of them were strong, and 0% [AK252] of V and A shapes. SES algorithm showed 0% [AH254] of rising parts, 0% [AJ254] of falling parts, 0% [AL254] of them were strong, and 0% [AK254] of V and A shapes.

IDP For knn, SC algorithm showed 17% [AH255] of rising parts, 17% [AJ255] of falling parts, 0% [AL255] of them were strong, and 100% [AK255] of V and A shapes. SDS algorithm showed 0% [AH257] of rising parts, 58% [AJ257] of falling parts, 75% [AL257] of them were strong, and 100% [AK257] of V and A shapes. SES algorithm showed 83% [AH259] of rising parts, 0% [AJ259] of falling parts, 75% [AL259] of them were strong, and 50% [AK259] of V and A shapes.
For pp/sigma, SC algorithm showed 0% [AH256] of rising

parts, 0% [AJ256] of falling parts, 0% [AL256] of them were strong, and 50% [AK256] of V and A shapes. SDS algorithm showed 0% [AH258] of rising parts, 33% [AJ258] of falling parts, 100% [AL258]of them were strong, and 50% [AK258] of V and A shapes. SES algorithm showed 0% [AH260] of rising parts, 0% [AJ260] of falling parts, 0% [AL260] of them were strong, and 0% [AK260] of V and A shapes.

Total For knn, SC algorithm showed 17% [AH261] of rising parts, 13% [AJ261] of falling parts, 25% [AL261] of them were strong, and 88% [AK261] of V and A shapes. SDS algorithm showed 4% [AH263] of rising parts, 50% [AJ263] of falling parts, 63% [AL263] of them were strong, and 88% [AK263] of V and A shapes. SES algorithm showed 75% [AH265] of rising parts, 0% [AJ265] of falling parts, 75% [AL265] of them were strong, and 25% [AK265] of V and A shapes.
For pp/sigma, SC algorithm showed 0% [AH262] of rising parts, 0% [AJ262] of falling parts, 0% [AL262] of them were strong, and 25% [AK262] of V and A shapes. SDS algorithm showed 0% [AH264] of rising parts, 25% [AJ264] of falling parts, 63% [AL264] of them were strong, and 25% [AK264] of V and A shapes. SES algorithm showed 0% [AH266] of rising parts, 0% [AJ266] of falling parts, 0% [AL266] of them were strong, and 0% [AK266] of V and A shapes.

6.2 Affinity

6.2.1 *Width.*

EN For knn, SC algorithm showed 19% [AC99] of narrow shapes, 47% [AE99] of wide shapes, and 42% [AF99] of changes. SDS algorithm showed 81% [AC101] of narrow shapes, 8% [AE101] of wide shapes, and 25% [AF101] of changes. SES algorithm showed 64% [AC103] of narrow shapes, 25% [AE103] of wide shapes, and 50% [AF103] of changes.
For pp, SC algorithm showed 81% [AC100] of narrow shapes, 0% [AE100] of wide shapes, and 50% [AF100] of changes. SDS algorithm showed 67% [AC102] of narrow shapes, 6% [AE102] of wide shapes, and 67% [AF102] of changes. SES algorithm showed 56% [AC104] of narrow shapes, 17% [AE104] of wide shapes, and 50% [AF104] of changes.

PL For knn, SC algorithm showed 29% [AG99] of narrow shapes, 58% [AI99] of wide shapes, and 88% [AJ99] of changes. SDS algorithm showed 8% [AG101] of narrow shapes, 71% [AI101] of wide shapes, and 50% [AJ101] of changes. SES algorithm showed 50% [AG103] of narrow shapes, 42% [AI103] of wide shapes, and 63% [AJ103] of changes.
For sigma, SC algorithm showed 8% [AG100] of narrow shapes, 71% [AI100] of wide shapes, and 38% [AJ100] of changes. SDS algorithm showed 0% [AG102] of narrow shapes, 75% [AI102] of wide shapes, and 13% [AJ102] of changes. SES algorithm showed 12.5% [AG104] of narrow shapes, 87.5% [AI104] of wide shapes, and 0% [AJ104] of changes.

Total For knn, SC algorithm showed 23% [AK99] of narrow shapes, 52% [AM99] of wide shapes, and 60% [AN99] of changes. SDS algorithm showed 52% [AK101] of narrow shapes, 33% [AM101] of wide shapes, and 35% [AN101] of changes. SES algorithm showed 58% [AK103] of narrow shapes, 32% [AM103] of wide shapes, and 55% [AN103] of changes.
For pp/sigma, SC algorithm showed 52% [AK100] of narrow shapes, 28% [AM100] of wide shapes, and 45% [AN100] of changes. SDS algorithm showed 40% [AK102] of narrow shapes, 33% [AM102] of wide shapes, and 45% [AN102] of changes. SES algorithm showed 38% [AK104] of narrow shapes, 45% [AM104] of wide shapes, and 30% [AN104] of changes.

6.2.2 *Shape.*

EN For knn, SC algorithm showed 3% [AC261] of rising parts, 6% [AE261] of falling parts, 0% [AG261] of them were strong, and 8% [AF261] of V and A shapes. SDS algorithm showed 8% [AC263] of rising parts, 8% [AE263] of falling parts, 8% [AG263] of them were strong, and 25% [AF263] of V and A shapes. SES algorithm showed 33% [AC265] of rising parts, 11% [AE265] of falling parts, 25% [AG265] of them were strong, and 33% [AF265] of V and A shapes.
For pp, SC algorithm showed 11% [AC262] of rising parts, 28% [AE262] of falling parts, 17% [AG262] of them were strong, and 33% [AF262] of V and A shapes. SDS algorithm showed 8% [AC264] of rising parts, 8% [AE264] of falling parts, 0% [AG264] of them were strong, and 17% [AF264] of V and A shapes. SES algorithm showed 14% [AC266] of rising parts, 19% [AE266] of falling parts, 8% [AG266] of them were strong, and 17% [AF266] of V and A shapes.

PL For knn, SC algorithm showed 17% [AH261] of rising parts, 13% [AJ261] of falling parts, 25% [AL261] of them were strong, and 88% [AK261] of V and A shapes. SDS algorithm showed 4% [AH263] of rising parts, 50% [AJ263] of falling parts, 63% [AL263] of them were strong, and 88% [AK263] of V and A shapes. SES algorithm showed 75% [AH265] of rising parts, 0% [AJ265] of falling parts, 75% [AL265] of them were strong, and 25% [AK265] of V and A shapes.
For sigma, SC algorithm showed 0% [AH262] of rising parts, 0% [AJ262] of falling parts, 0% [AL262] of them were strong, and 25% [AK262] of V and A shapes. SDS algorithm showed 0% [AH264] of rising parts, 25% [AJ264] of falling parts, 63% [AL264] of them were strong, and 25% [AK264] of V and A shapes. SES algorithm showed 0% [AH266] of rising parts, 0% [AJ266] of falling parts, 0% [AL266] of them were strong, and 0% [AK266] of V and A

Total For knn, SC algorithm showed 8% [AM261] of rising parts, 8% [AO261] of falling parts, 10% [AQ261] of them were strong, and 40% [AP261] of V and A shapes. SDS algorithm showed 7% [AM263] of rising parts, 25% [AO263] of falling parts, 30% [AQ263] of them were strong, and 50% [AP263] of V and A shapes. SES algorithm showed 50% [AM265] of rising parts, 7% [AO265] of falling parts, 45% [AQ265] of them were strong, and 30% [AP265] of V and A shapes.
For pp/sigma, SC algorithm showed 7% [AM262] of rising parts, 17% [AO262] of falling parts, 10% [AQ262] of them were strong, and 30% [AP262] of V and A shapes. SDS algorithm showed 5% [AM264] of rising parts, 15% [AO264] of falling parts, 25% [AQ264] of them were strong, and 20% [AP264] of V and A shapes. SES algorithm showed 8% [AM266]

of rising parts, 12% [AO266] of falling parts, 5% [AQ266] of them were strong, and 10% [AP266] of V and A

6.3 Sparsity

6.3.1 Width.

DN For knn, SC algorithm showed 8% [AC127] of narrow shapes, 42% [AE127] of wide shapes, and 50% [AF127] of changes. SDS algorithm showed 58% [AC129] of narrow shapes, 17% [AE129] of wide shapes, and 50% [AF129] of changes. SES algorithm showed 33% [AC131] of narrow shapes, 67% [AE131] of wide shapes, and 100% [AF131] of changes.
For pp, SC algorithm showed 92% [AC128] of narrow shapes, 0% [AE128] of wide shapes, and 25% [AF128] of changes. SDS algorithm showed 50% [AC130] of narrow shapes, 8% [AE130] of wide shapes, and 75% [AF130] of changes. SES algorithm showed 17% [AC132] of narrow shapes, 50% [AE132] of wide shapes, and 100% [AF132] of changes.

SP For knn, SC algorithm showed 29% [AG127] of narrow shapes, 54% [AI127] of wide shapes, and 63% [AJ127] of changes. SDS algorithm showed 50% [AG129] of narrow shapes, 42% [AI129] of wide shapes, and 13% [AJ129] of changes. SES algorithm showed 50% [AG131] of narrow shapes, 33% [AI131] of wide shapes, and 50% [AJ131] of changes.
For pp/sigma, SC algorithm showed 50% [AG128] of narrow shapes, 21% [AI128] of wide shapes, and 50% [AJ128] of changes. SDS algorithm showed 38% [AG130] of narrow shapes, 38% [AI129] of wide shapes, and 50% [AJ130] of changes. SES algorithm showed 42% [AG132] of narrow shapes, 38% [AI132] of wide shapes, and 13% [AJ132] of changes.

SS For knn, SC algorithm showed 22% [AK127] of narrow shapes, 50% [AM127] of wide shapes, and 58% [AN127] of changes. SDS algorithm showed 53% [AK129] of narrow shapes, 33% [AM129] of wide shapes, and 42% [AN129] of changes. SES algorithm showed 44% [AK131] of narrow shapes, 44% [AM131] of wide shapes, and 67% [AN131] of changes.
For pp/sigma, SC algorithm showed 64% [AK128] of narrow shapes, 14% [AM128] of wide shapes, and 42% [AN128] of changes. SDS algorithm showed 42% [AK130] of narrow shapes, 28% [AM129] of wide shapes, and 67% [AN130] of changes. SES algorithm showed 33% [AK132] of narrow shapes, 42% [AM132] of wide shapes, and 42% [AN132] of changes.

Total For knn, SC algorithm showed 25% [AO127] of narrow shapes, 52% [AQ127] of wide shapes, and 60% [AR127] of changes. SDS algorithm showed 52% [AO129] of narrow shapes, 37% [AQ129] of wide shapes, and 30% [AR129] of changes. SES algorithm showed 47% [AO131] of narrow shapes, 40% [AQ131] of wide shapes, and 60% [AR131] of changes.
For pp/sigma, SC algorithm showed 58% [AO128] of narrow shapes, 17% [AQ128] of wide shapes, and 45% [AR128] of changes. SDS algorithm showed 40% [AO130] of narrow shapes, 32% [AQ129] of wide shapes, and 60% [AR130] of changes. SES algorithm showed 37% [AO132] of narrow shapes, 40% [AQ132] of wide shapes, and 30% [AR132] of changes.

6.3.2 Shape.

DN For knn, SC algorithm showed 8% [AC289] of rising parts, 8% [AE289] of falling parts, 0% [AG289] of them were strong, and 0% [AF289] of V and A shapes. SDS algorithm showed 17% [AC291] of rising parts, 8% [AE291] of falling parts, 0% [AG291] of them were strong, and 50% [AF291] of V and A shapes. SES algorithm showed 50% [AC293] of rising parts, 25% [AE293] of falling parts, 75% [AG293] of them were strong, and 100% [AF293] of V and A shapes.
For pp, SC algorithm showed 8% [AC290] of rising parts, 42% [AE290] of falling parts, 25% [AG290] of them were strong, and 0% [AF290] of V and A shapes. SDS algorithm showed 17% [AC292] of rising parts, 8% [AE292] of falling parts, 0% [AG292] of them were strong, and 25% [AF292] of V and A shapes. SES algorithm showed 8% [AC294] of rising parts, 50% [AE294] of falling parts, 25% [AG294] of them were strong, and 25% [AF294] of V and A shapes.

SP For knn, SC algorithm showed 4% [AH289] of rising parts, 8% [AJ289] of falling parts, 0% [AL289] of them were strong, and 50% [AK289] of V and A shapes. SDS algorithm showed 8% [AH291] of rising parts, 29% [AJ291] of falling parts, 25% [AL291] of them were strong, and 63% [AK291] of V and A shapes. SES algorithm showed 46% [AH293] of rising parts, 4% [AJ293] of falling parts, 25% [AL293] of them were strong, and 13% [AK293] of V and A shapes.
For pp/sigma, SC algorithm showed 0% [AH290] of rising parts, 17% [AJ290] of falling parts, 13% [AL290] of them were strong, and 25% [AK290] of V and A shapes. SDS algorithm showed 4% [AH292] of rising parts, 17% [AJ292] of falling parts, 38% [AL292] of them were strong, and 25% [AK292] of V and A shapes. SES algorithm showed 8% [AH294] of rising parts, 4% [AJ294] of falling parts, 0% [AL294] of them were strong, and 13% [AK294] of V and A shapes.

SS For knn, SC algorithm showed 13% [AM289] of rising parts, 8% [AO289] of falling parts, 25% [AQ289] of them were strong, and 50% [AP289] of V and A shapes. SDS algorithm showed 0% [AM291] of rising parts, 29% [AO291] of falling parts, 50% [AQ291] of them were strong, and 38% [AP291] of V and A shapes. SES algorithm showed 54% [AM293] of rising parts, 0% [AO293] of falling parts, 50% [AQ293] of them were strong, and 13% [AP293] of V and A shapes.
For pp/sigma, SC algorithm showed 13% [AM290] of rising parts, 4% [AO290] of falling parts, 0% [AQ290] of them were strong, and 50% [AP290] of V and A shapes. SDS algorithm showed 0% [AM292] of rising parts, 17% [AO292] of falling parts, 25% [AQ292] of them were strong, and 13% [AP292] of V and A shapes. SES algorithm showed 8% [AM294] of rising parts, 0% [AO294] of falling parts, 0% [AQ294] of them were strong, and 0% [AP294] of V and A shapes.

Total For knn, SC algorithm showed 8% [AR289] of rising parts, 8% [AT289] of falling parts, 10% [AV289] of them were strong, and 40% [AU289] of V and A shapes. SDS algorithm showed 7% [AR291] of rising parts, 25% [AT291] of falling parts, 30% [AV291] of them were strong, and 50% [AU291] of V and A shapes. SES algorithm showed 50% [AR293] of rising parts, 7% [AT293] of falling parts, 45% [AV293] of them were strong,

and 30% [AU293] of V and A shapes.
For pp/sigma, SC algorithm showed 7% [AR290] of rising parts, 17% [AT290] of falling parts, 10% [AV290] of them were strong, and 30% [AU290] of V and A shapes. SDS algorithm showed 5% [AR292] of rising parts, 15% [AT292] of falling parts, 25% [AV292] of them were strong, and 20% [AU292] of V and A shapes. SES algorithm showed 8% [AR294] of rising parts, 12% [AT294] of falling parts, 5% [AV294] of them were strong, and 10% [AU294] of V and A shapes.

7 SPARSITY TYPE

7.1 Protein type

7.1.1 *Width.*

NFP For knn, dense data showed 33% [AC133] of narrow shapes, 39% [AE133] of wide shapes, and 50% [AF133] of changes. sparse data showed 53% [AG133] of narrow shapes, 36% [AI133] of wide shapes, and 33% [AJ133] of changes. supersparse data showed 46% [AK133] of narrow shapes, 37% [AM133] of wide shapes, and 44% [AN133] of changes.
For pp/sigma, dense data showed 56% [AC134] of narrow shapes, 17% [AE134] of wide shapes, and 50% [AF134] of changes. sparse data showed 50% [AG134] of narrow shapes, 25% [AI134] of wide shapes, and 33% [AJ134] of changes. supersparse data showed 52% [AK134] of narrow shapes, 22% [AM134] of wide shapes, and 44% [AN134] of changes.

IDP For knn, dense data showed 33% [AC135] of narrow shapes, 44% [AE135] of wide shapes, and 83% [AF135] of changes. sparse data showed 33% [AG135] of narrow shapes, 50% [AI135] of wide shapes, and 50% [AJ135] of changes. supersparse data showed 33% [AK135] of narrow shapes, 48% [AM135] of wide shapes, and 67% [AN135] of changes.
For pp/sigma, dense data showed 50% [AC136] of narrow shapes, 22% [AE136] of wide shapes, and 83% [AF136] of changes. sparse data showed 36% [AG136] of narrow shapes, 39% [AI136] of wide shapes, and 42% [AJ136] of changes. supersparse data showed 41% [AK136] of narrow shapes, 33% [AM136] of wide shapes, and 56% [AN136] of changes.

Total For knn, dense data showed 33% [AC137] of narrow shapes, 42% [AE137] of wide shapes, and 67% [AF137] of changes. sparse data showed 43% [AG137] of narrow shapes, 43% [AI137] of wide shapes, and 42% [AJ137] of changes. supersparse data showed 40% [AK137] of narrow shapes, 43% [AM137] of wide shapes, and 56% [AN137] of changes.
For pp/sigma, dense data showed 53% [AC138] of narrow shapes, 19% [AE138] of wide shapes, and 67% [AF138] of changes. sparse data showed 43% [AG138] of narrow shapes, 32% [AI138] of wide shapes, and 38% [AJ138] of changes. supersparse data showed 46% [AK138] of narrow shapes, 28% [AM138] of wide shapes, and 50% [AN138] of changes.

7.1.2 *Shape.*

NFP For knn, dense data showed 28% [AC295] of rising parts, 17% [AE295] of falling parts, 17% [AG295] of them were strong, and 50% [AF295] of V and A shapes. sparse data showed 17% [AH295] of rising parts, 8% [AJ295] of falling parts, 17% [AL295] of them were strong, and 25% [AK295] of V and A shapes. supersparse data showed 19% [AM295] of rising parts, 8% [AO295] of falling parts, 42% [AQ295] of them were strong, and 25% [AP295] of V and A shapes.
For pp/sigma, dense data showed 11% [AC296] of rising parts, 17% [AE296] of falling parts, 0% [AG296] of them were strong, and 17% [AF296] of V and A shapes. sparse data showed 6% [AH296] of rising parts, 11% [AJ296] of falling parts, 8% [AL296] of them were strong, and 17% [AK296] of V and A shapes. supersparse data showed 3% [AM296] of rising parts, 3% [AO296] of falling parts, 0% [AQ296] of them were strong, and 8% [AP296] of V and A shapes.

IDP For knn, dense data showed 22% [AC297] of rising parts, 11% [AE297] of falling parts, 33% [AG297] of them were strong, and 50% [AF297] of V and A shapes. sparse data showed 22% [AH297] of rising parts, 19% [AJ297] of falling parts, 17% [AL297] of them were strong, and 58% [AK297] of V and A shapes. supersparse data showed 25% [AM297] of rising parts, 17% [AO297] of falling parts, 42% [AQ297] of them were strong, and 42% [AP297] of V and A shapes.
For pp/sigma, dense data showed 11% [AC298] of rising parts, 50% [AE298] of falling parts, 33% [AG298] of them were strong, and 17% [AF298] of V and A shapes. sparse data showed 3% [AH298] of rising parts, 14% [AJ298] of falling parts, 25% [AL298] of them were strong, and 25% [AK298] of V and A shapes. supersparse data showed 11% [AM298] of rising parts, 11% [AO298] of falling parts, 17% [AQ298] of them were strong, and 33% [AP298] of V and A shapes.

Total For knn, dense data showed 25% [AC299] of rising parts, 14% [AE299] of falling parts, 25% [AG299] of them were strong, and 50% [AF299] of V and A shapes. sparse data showed 19% [AH299] of rising parts, 14% [AJ299] of falling parts, 17% [AL299] of them were strong, and 42% [AK299] of V and A shapes. supersparse data showed 22% [AM299] of rising parts, 13% [AO299] of falling parts, 42% [AQ299] of them were strong, and 33% [AP299] of V and A shapes.
For pp/sigma, dense data showed 11% [AC300] of rising parts, 33% [AE300] of falling parts, 17% [AG300] of them were strong, and 17% [AF300] of V and A shapes. sparse data showed 4% [AH300] of rising parts, 13% [AJ300] of falling parts, 17% [AL300] of them were strong, and 21% [AK300] of V and A shapes. supersparse data showed 7% [AM300] of rising parts, 7% [AO300] of falling parts, 8% [AQ300] of them were strong, and 21% [AP300] of V and A shapes.

7.2 Affinity type

7.2.1 *Width.*

EN For knn, dense data showed 33% [AK5] of narrow shapes, 42% [AM5] of wide shapes, and 67% [AN5] of changes. sparse data showed 69% [AK7] of narrow shapes, 8% [AM7] of wide shapes, and 25% [AN7] of changes. supersparse data showed 61% [AK9] of narrow shapes, 31% [AM9] of wide shapes, and 25% [AN9] of changes.
For pp, dense data showed 53% [AK6] of narrow shapes, 19% [AM6] of wide shapes, and 67% [AN6] of changes. sparse data showed 72% [AK8] of narrow shapes, 0% [AM8] of wide shapes, and 50% [AN8] of changes. supersparse data showed

78% [AK10] of narrow shapes, 3% [AM10] of wide shapes, and 50% [AN10] of changes.

PL For knn, sparse data showed 17% [AK13] of narrow shapes, 78% [AM13] of wide shapes, and 58% [AN13] of changes. supersparse data showed 42% [AK15] of narrow shapes, 36% [AM15] of wide shapes, and 75% [AN15] of changes.
For sigma, sparse data showed 14% [AK14] of narrow shapes, 64% [AM14] of wide shapes, and 25% [AN14] of changes. supersparse data showed 0% [AK16] of narrow shapes, 92% [AM16] of wide shapes, and 8% [AN16] of changes.

Total For knn, dense data showed 33% [AK17] of narrow shapes, 42% [AM17] of wide shapes, and 67% [AN17] of changes. sparse data showed 43% [AK19] of narrow shapes, 43% [AM19] of wide shapes, and 42% [AN19] of changes. supersparse data showed 51% [AK21] of narrow shapes, 33% [AM21] of wide shapes, and 50% [AN21] of changes.
For pp/sigma, dense data showed 53% [AK18] of narrow shapes, 19% [AM18] of wide shapes, and 67% [AN18] of changes. sparse data showed 43% [AK20] of narrow shapes, 32% [AM20] of wide shapes, and 38% [AN20] of changes. supersparse data showed 39% [AK22] of narrow shapes, 47% [AM22] of wide shapes, and 29% [AN22] of changes.

7.2.2 *Shape.*

EN For knn, dense data showed 25% [AM145] of rising parts, 14% [AO145] of falling parts, 25% [AQ145] of them were strong, and 50% [AP145] of V and A shapes. sparse data showed 14% [AM147] of rising parts, 8% [AO147] of falling parts, 0% [AQ147] of them were strong, and 17% [AP147] of V and A shapes. supersparse data showed 6% [AM149] of rising parts, 3% [AO149] of falling parts, 8% [AQ149] of them were strong, and 0% [AP149] of V and A shapes.
For pp, dense data showed 11% [AM146] of rising parts, 33% [AO146] of falling parts, 17% [AQ146] of them were strong, and 17% [AP146] of V and A shapes. sparse data showed 8% [AM148] of rising parts, 17% [AO148] of falling parts, 8% [AQ148] of them were strong, and 25% [AP148] of V and A shapes. supersparse data showed 14% [AM150] of rising parts, 6% [AO150] of falling parts, 0% [AQ150] of them were strong, and 25% [AP150] of V and A shapes.

PL For knn, sparse data showed 25% [AM153] of rising parts, 19% [AO153] of falling parts, 33% [AQ153] of them were strong, and 67% [AP153] of V and A shapes. supersparse data showed 39% [AM155] of rising parts, 22% [AO155] of falling parts, 75% [AQ155] of them were strong, and 67% [AP155] of V and A shapes.
For sigma, sparse data showed 0% [AM154] of rising parts, 8% [AO154] of falling parts, 25% [AQ154] of them were strong, and 17% [AP154] of V and A shapes. supersparse data showed 0% [AM156] of rising parts, 8% [AO156] of falling parts, 17% [AQ156] of them were strong, and 17% [AP156] of V and A shapes.

Total For knn, dense data showed 25% [AM157] of rising parts, 14% [AO157] of falling parts, 25% [AQ157] of them were strong, and 50% [AP157] of V and A shapes. sparse data showed 19% [AM159] of rising parts, 14% [AO159] of falling parts, 17% [AQ159] of them were strong, and 42% [AP159] of V and A shapes. supersparse data showed 22% [AM161] of rising parts, 13% [AO161] of falling parts, 42% [AQ161] of them were strong, and 33% [AP161] of V and A shapes.
For pp/sigma, dense data showed 11% [AM158] of rising parts, 33% [AO158] of falling parts, 17% [AQ158] of them were strong, and 17% [AP158] of V and A shapes. sparse data showed 4% [AM160] of rising parts, 13% [AO160] of falling parts, 17% [AQ160] of them were strong, and 21% [AP160] of V and A shapes. supersparse data showed 7% [AM162] of rising parts, 7% [AO162] of falling parts, 8% [AQ162] of them were strong, and 21% [AP162] of V and A shapes.

7.3 Algorithms

7.3.1 *Width.*

SC For knn, dense data showed 8% [AC127] of narrow shapes, 42% [AE127] of wide shapes, and 50% [AF127] of changes. sparse data showed 29% [AG127] of narrow shapes, 54% [AI127] of wide shapes, and 63% [AJ127] of changes. supersparse data showed 22% [AK127] of narrow shapes, 50% [AM127] of wide shapes, and 58% [AN127] of changes.
For pp/sigma, dense data showed 92% [AC128] of narrow shapes, 0% [AE128] of wide shapes, and 25% [AF128] of changes. sparse data showed 50% [AG128] of narrow shapes, 21% [AI128] of wide shapes, and 50% [AJ128] of changes. supersparse data showed 64% [AK128] of narrow shapes, 14% [AM128] of wide shapes, and 42% [AN128] of changes.

SDS For knn, dense data showed 58% [AC129] of narrow shapes, 17% [AE129] of wide shapes, and 50% [AF129] of changes. sparse data showed 50% [AG129] of narrow shapes, 42% [AI129] of wide shapes, and 13% [AJ129] of changes. supersparse data showed 53% [AK129] of narrow shapes, 33% [AM129] of wide shapes, and 42% [AN129] of changes.
For pp/sigma, dense data showed 50% [AC130] of narrow shapes, 8% [AE130] of wide shapes, and 75% [AF130] of changes. sparse data showed 38% [AG130] of narrow shapes, 38% [AI130] of wide shapes, and 50% [AJ130] of changes. supersparse data showed 42% [AK130] of narrow shapes, 28% [AM130] of wide shapes, and 67% [AN130] of changes.

SES For knn, dense data showed 33% [AC131] of narrow shapes, 67% [AE131] of wide shapes, and 100% [AF131] of changes. sparse data showed 50% [AG131] of narrow shapes, 33% [AI131] of wide shapes, and 50% [AJ131] of changes. supersparse data showed 44% [AK131] of narrow shapes, 44% [AM131] of wide shapes, and 67% [AN131] of changes. For pp/sigma, dense data showed 17% [AC132] of narrow shapes, 50% [AE132] of wide shapes, and 100% [AF132] of changes.
sparse data showed 42% [AG132] of narrow shapes, 38% [AI132] of wide shapes, and 13% [AJ132] of changes. supersparse data showed 33% [AK132] of narrow shapes, 42% [AM132] of wide shapes, and 42% [AN132] of changes.

Total For knn, dense data showed 33% [AC137] of narrow shapes, 42% [AE137] of wide shapes, and 83% [AF137] of changes. sparse data showed 43% [AG137] of narrow shapes, 43% [AI137] of wide shapes, and 42% [AJ137] of changes. supersparse data showed 40% [AK137] of narrow shapes, 43%

[AM137] of wide shapes, and 56% [AN137] of changes.
For pp/sigma, dense data showed 53% [AC138] of narrow shapes, 19% [AE138] of wide shapes, and 75% [AF138] of changes. sparse data showed 43% [AG138] of narrow shapes, 32% [AI138] of wide shapes, and 38% [AJ138] of changes. supersparse data showed 46% [AK138] of narrow shapes, 28% [AM138] of wide shapes, and 50% [AN138] of changes.

7.3.2 *Shape.*

SC For knn, dense data showed 8% [AC289] of rising parts, 8% [AE289] of falling parts, 0% [AQ289] of them were strong, and 0% [AP289] of V and A shapes. sparse data showed 4% [AH289] of rising parts, 8% [AJ289] of falling parts, 0% [AL289] of them were strong, and 50% [AK289] of V and A shapes. supersparse data showed 13% [AM289] of rising parts, 8% [AO289] of falling parts, 25% [AQ289] of them were strong, and 50% [AP289] of V and A shapes.
For pp/sigma, dense data showed 8% [AC290] of rising parts, 42% [AE290] of falling parts, 25% [AQ290] of them were strong, and 0% [AP290] of V and A shapes. sparse data showed 0% [AH290] of rising parts, 17% [AJ290] of falling parts, 13% [AL290] of them were strong, and 25% [AK290] of V and A shapes. supersparse data showed 13% [AM290] of rising parts, 4% [AO290] of falling parts, 0% [AQ290] of them were strong, and 50% [AP290] of V and A shapes.

SDS For knn, dense data showed 17% [AC291] of rising parts, 8% [AE291] of falling parts, 0% [AQ291] of them were strong, and 50% [AP291] of V and A shapes. sparse data showed 8% [AH291] of rising parts, 29% [AJ291] of falling parts, 25% [AL291] of them were strong, and 63% [AK291] of V and A shapes. supersparse data showed 0% [AM291] of rising parts, 29% [AO291] of falling parts, 50% [AQ291] of them were strong, and 38% [AP291] of V and A shapes.
For pp/sigma, dense data showed 17% [AC292] of rising parts, 8% [AE292] of falling parts, 0% [AQ292] of them were strong, and 25% [AP292] of V and A shapes. sparse data showed 4% [AH292] of rising parts, 17% [AJ292] of falling parts, 38% [AL292] of them were strong, and 25% [AK292] of V and A shapes. supersparse data showed 0% [AM292] of rising parts, 17% [AO292] of falling parts, 25% [AQ292] of them were strong, and 13% [AP292] of V and A shapes.

SES For knn, dense data showed 50% [AC291] of rising parts, 25% [AE291] of falling parts, 75% [AQ291] of them were strong, and 100% [AP291] of V and A shapes. sparse data showed 46% [AH291] of rising parts, 4% [AJ291] of falling parts, 25% [AL291] of them were strong, and 13% [AK291] of V and A shapes. supersparse data showed 54% [AM291] of rising parts, 0% [AO291] of falling parts, 50% [AQ291] of them were strong, and 13% [AP291] of V and A shapes.
For pp/sigma, dense data showed 8% [AC292] of rising parts, 50% [AE292] of falling parts, 25% [AQ292] of them were strong, and 25% [AP292] of V and A shapes. sparse data showed 8% [AH292] of rising parts, 4% [AJ292] of falling parts, 0% [AL292] of them were strong, and 13% [AK292] of V and A shapes. supersparse data showed 8% [AM292] of rising parts, 0% [AO292] of falling parts, 0% [AQ292] of them were strong, and 0% [AP292] of V and A shapes.

Total For knn, dense data showed 25% [AC299] of rising parts, 14% [AE299] of falling parts, 25% [AQ299] of them were strong, and 50% [AP299] of V and A shapes. sparse data showed 19% [AH299] of rising parts, 14% [AJ299] of falling parts, 17% [AL299] of them were strong, and 42% [AK299] of V and A shapes. supersparse data showed 22% [AM299] of rising parts, 13% [AO299] of falling parts, 42% [AQ299] of them were strong, and 33% [AP299] of V and A shapes.
For pp/sigma, dense data showed 11% [AC300] of rising parts, 33% [AE300] of falling parts, 17% [AQ300] of them were strong, and 17% [AP300] of V and A shapes. sparse data showed 4% [AH300] of rising parts, 13% [AJ300] of falling parts, 17% [AL300] of them were strong, and 21% [AK300] of V and A shapes. supersparse data showed 7% [AM300] of rising parts, 7% [AO300] of falling parts, 8% [AQ300] of them were strong, and 21% [AP300] of V and A shapes.

REFERENCES

[1] Stephen Boyd and Lieven Vandenberghe. 2004. *Convex Optimization.* Cambridge University Press, New York, NY, USA.

[2] Ehsan Elhamifar and René Vidal. 2012. Sparse Subspace Clustering: Algorithm, Theory, and Applications. *CoRR* abs/1203.1005 (2012). http://arxiv.org/abs/1203.1005

[3] Geoffrey E Hinton and Sam T Roweis. 2003. Stochastic neighbor embedding. In *Advances in neural information processing systems*. 857–864.

[4] Max Vladymyrov and Miguel Carreira-perpinan. 2013. Entropic Affinities: Properties and Efficient Numerical Computation. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, Sanjoy Dasgupta and David Mcallester (Eds.), Vol. 28. JMLR Workshop and Conference Proceedings, 477–485. http://jmlr.org/proceedings/papers/v28/vladymyrov13.pdf

[5] Ulrike Von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and computing* 17, 4 (2007), 395–416.