Chapter 7

Data Clustering in Non-stationary Environments using a Local Network Neighbourhood Artificial Immune System

A non-stationary environment can be defined as feature vectors in space which move or adapt to different spatial positions over time [64]. The data is thus dynamic over time. Clustering of non-stationary data results into different partitions of the data at different points in time and depends on the severity and the frequency of change in the data. Therefore, from a clustering perspective in a non-stationary environment, the initial formed clusters of a data set can *adapt* over time. This means that, at each time step, the feature vectors associated with different clusters can follow different *migration* types to and from other clusters. The *migration* of feature vectors from one cluster to another implies that the centroids of the different clusters can also move in space to different positions. Therefore, clusters (centroids) may move, disappear and/or new clusters may appear.

This chapter investigates different data migration types and proposes a technique to generate artificial non-stationary data which follows different migration types. Furthermore, the chapter revises the proposed clustering performance measures in section 2.5 which are more applicable to measure the clustering quality in a non-stationary environment compared to the clustering performance measures for stationary environments. The proposed clustering performance measures are then used to compare the clustering results of LNNAIS and LNN_{SDOT} with two other network based artificial immune models.

Section 7.1 revises the clustering performance measures as discussed in section 2.5 that can be used to evaluate the partitioning quality of clustering algorithms in non-stationary environments. This is followed by a discussion and investigation into different data *migration* types in section 7.2. Section 7.3 proposes a technique to generate artificial non-stationary data with different migration types. Section 7.4 analyses and discusses the sensitivity of the LNNAIS parameters on the different artificial non-stationary data sets for each of the defined data migration types. Section 7.5 discusses and compares the clustering results obtained by LNNAIS, LNN_{SDOT}, SMAIN and DWB to cluster generated artificial non-stationary data in different dimensions with different cluster sizes, frequencies of change and severities of change.

7.1 Clustering Performance Measures for Non-stationary Environments

Clustering of a data set at a specific point in time, t, is the partitioning of the data set such that patterns within the same partition are more *similar* when compared to patterns which form part of different partitions. The partitioning of a data set into different clusters may differ among different clustering algorithms at a specific point in time. Also, clusters may differ among different points in time. Therefore the clustering quality needs to be evaluated at each point in time. The quality of the clusters can be validated using a cluster validity index. The cluster validity index used in this chapter was proposed by Ray and Turi [151] (as defined in equation (2.49)).

In the context of clustering of non-stationary environments, J_{intra} (cluster compactness) and J_{inter} (cluster separation) can be used, in addition to the validity index, Q_{ratio}, to quantify the quality of partitioning by clustering algorithms over time. An example of a clustering algorithm's partitioning in a non-stationary environment was given in section 2.5. The average separation (J_{inter}) plotted against time will increase in value if the clusters become more separated in time, i.e. clusters move away from one another. If there is any *migration* of feature vectors between clusters, it is expected that the average *intra-clustering* distance (J_{intra}) plotted against time will fluctuate from the time of *migration* until the feature vectors become stationary again (data *migration* types are discussed in section 7.2). It is expected that a change in the number of clusters will result in a change in J_{intra} and/or J_{inter}.

The measured clustering quality (Q_{ratio}) of different clustering algorithms can be compared by averaging each algorithm's clustering quality measure at each step in time over the total running time T (as proposed in section 2.8). This will give each algorithm a mean value of measured clustering quality, $\bar{Q}_i$, for a specific run, i, calculated as

$$\bar{Q}_i = \frac{\sum_{t=1}^{T} Q_{best}(t)}{T} \tag{7.1}$$

where $Q_{best}(t)$ is the cluster quality at time t, calculated as

$$Q_{best}(t) = \min\left\{Q_{ratio}\left(K^{(t)}\right) : 1 < K^{(t)} \leq |P^{(t)}|\right\} \tag{7.2}$$

and Q_{ratio} is minimised by an optimal partitioning of data set $P^{(t)}$ into $K^{(t)}$ clusters at time t. The trajectory of the clustering quality across the entire dynamic landscape is then calculated by averaging $\bar{Q}_i$ over the number of independent runs E, referred to as the collective mean quality. The collective mean quality, $\hat{Q}$, is calculated as

$$\hat{Q} = \frac{\sum_{i=1}^{E} \bar{Q}_i}{E} \tag{7.3}$$

$\hat{Q}$ is derived from the collective mean fitness which is defined in [131] for function optimisation in non-stationary environments. The collective mean fitness takes into account the fitness trajectory across the entire dynamic landscape [131] by averaging the mean value of measured performance over the number of runs E.

The next section discusses the different data *migration* types that can occur in non-stationary environments, as investigated in this chapter.

7.2 Data Migration Types

Feature vectors in a data set can change at any point in time with different severities. Feature vectors in a non-stationary data set can follow different migration types. Based on the different migration types which were discussed in sections 2.4 and 2.5, three generic data migration types can be identified:

- *Pattern migration*: A feature vector can migrate from one cluster to another in the data set. The severity of change, $\tilde{s}$, can be expressed as a percentage of all the feature vectors among the different clusters in the data set which can each migrate to a randomly selected cluster in the data set. Pattern migration can result in some clusters to decrease in size while other clusters increase in size.

- *Cluster migration*: Similar to pattern migration, but instead of selecting a ratio of feature vectors for migration, a fraction of the number of clusters in the data set is selected for migration. Each pattern of a selected cluster migrates to a randomly selected cluster (which was not selected for migration). Cluster migration will result in the disappearance of clusters. In the case where the patterns of the selected cluster migrate to random spatial positions, new clusters will appear.

- *Centroid migration*: All the clusters in a data set adapt the spatial position of their centroids in such a way that feature vectors associated with each cluster remain part of that cluster after the change. The severity of change for centroid migration is discussed in more detail in section 7.3. Centroid migration can result in merging or division of clusters.

The number of times data changes occur within a fixed period of time is referred to as the *frequency* of change, $\tilde{f}$. A higher number of changes within a fixed period of time results in a higher frequency of change in the data and vice versa.

The next section discusses the procedure followed to generate artificial non-stationary data with different frequencies and severities of change in the data.

7.3 Generating Artificial Non-stationary Data

Referring to section 2.1, a cluster, C_k, is a partition of patterns and is represented by a centroid, $\mathbf{c}_k$. The distances between a centroid and the patterns of the cluster determine the *compactness* of the cluster. Therefore a cluster can be generated using the following multidimensional Gaussian function:

$$g(\mathbf{x}_k, \mathbf{c}_k) = a \exp^{-\left[\sum_{n=1}^{N} \frac{\left(x_{k,n} - c_{k,n}\right)^2}{2\sigma_{k,n}^2}\right]} \tag{7.4}$$

where a is the amplitude, $\mathbf{x}_{k,n}$ is the offset from the centroid $\mathbf{c}_{k,n}$ in dimension n, and $\sigma_{k,n}$ is the spread (compactness) in dimension n. Different types of clusters can be generated using these Gaussian function parameters differing in centroid position ($\mathbf{c}$), compactness (σ) and patterns ($\mathbf{c} \pm \mathbf{x}$) which is associated with the cluster.

In order to simulate a non-stationary environment, the generated Gaussian clusters need to follow a specific migration type (the different migration types were discussed in section 7.2). Focusing on *centroid migration*, the centroid of a cluster needs to change along a specific path. A hypersphere in N-dimensional space ((N-1)-dimensional sphere) with a radius, φ, and middle point, $\mathbf{m}$, was used as the path. A centroid, $\mathbf{c}_k$, is represented by an angle vector, θ_k. The angle vector is projected onto the surface of the (N-1)-dimensional sphere to determine the centroid $\mathbf{c}_k$. A change in $\theta_{k,n-1}$ will result in a projected change in $\mathbf{c}_{k,n}$ and move the cluster centroid on the surface of the (N-1)-dimensional sphere. The severity of change in an angle is a ratio, $\frac{\tilde{s}}{10}$, of the angle difference between $\theta_{k,n-1}$ and a randomly generated angle sampled from $U[0,\pi]$.

Let $\mathbf{c}^{'}_{k,n}$ be the new centroid position in dimension n after a change in $\theta_{k,n-1}$. The spread of patterns in cluster C_k (compactness) needs to remain the same as before the change. Therefore the offset (vector difference) between $\mathbf{c}^{'}_{k,n}$ and the previous centroid $\mathbf{c}_{k,n}$ needs to be added (vector addition) to each pattern in cluster C_k.

The remaining migration types are simulated by removing patterns from a cluster and generating new random patterns for the remaining clusters such that the initial total number of patterns, $|P|$, remains the same. The data in this chapter was generated in different dimensions ($N \in [3,8,15]$). Each data set initially consists of eight clusters ($K = 8$) which are uniformly distributed on the surface of the (N-1)-dimensional sphere ($\varphi = 15$ and $\mathbf{m}$ is zero in every dimension). Each cluster is generated with the Gaussian function of equation (7.4) where $a = 1$ and $\sigma = 1$ (for each dimension). Clusters initially have the same size. The clusters are also generated in different sizes of $[10,25,50]$ ($|P| = K \times |C| \in [80,200,400]$). Data was generated for each migration type at different frequencies, $\tilde{f} \in [1,2,3,4,5]$, and different severities of change, $\tilde{s} \in [1,2,3,4,5]$ giving 225 different non-stationary data sets for each migration type. The generated artificial non-stationary data sets represent a good distribution of data clustering problems in non-stationary environments with the number of features in the range $[3,8,15]$ and the number of patterns in the range $[80,200,400]$ which change/migrate at different frequencies $\tilde{f} \in [1,2,3,4,5]$ and severities $\tilde{s} \in [1,2,3,4,5]$ of change. The different values of $\tilde{f}$ and $\tilde{s}$ are expressed as a ratio $\frac{\tilde{f}}{10}$ and $\frac{\tilde{s}}{10}$ respectively. The ratio of $\tilde{f}$ is then multiplied with the total number of time steps T to determine the time step size at which a change occurs ($\frac{\tilde{f}}{10} \times T$). Therefore higher values of $\tilde{f}$ imply lower frequencies of change. The ratio of $\tilde{s}$ is used to determine the severity of change for the applicable migration type. Higher values of $\tilde{s}$ imply higher severities of change.

The following sections discuss the sensitivity of the LNNAIS parameters to changes in dimension, cluster size, frequency of change and severity of change in the data for each of the migration types. All experimental results reported in the following sections are collective means taken over 100 time steps ($T = 100$) for 50 runs ($E = 50$) unless stated otherwise. A time step is a single presentation of the data set at that specific point in time. The parameter values for each non-stationary data set were found empirically to deliver the best performance for each of the models.

7.4 Sensitivity of LNNAIS parameters

This section discusses the sensitivity of the LNNAIS parameters for each of the migration types. Parallel coordinates [43, 94] is used to illustrate the effect of changes in the non-stationary environment (e.g. number of dimensions, cluster size, frequency of change and severity of change) on the parameters of LNNAIS.

Parallel coordinates is a visualisation technique to analyse multivariate data. The attribute values of a pattern in an N dimensional data set are each plotted on a parallel line. Each parallel line represents a dimension, therefore an N dimensional data set has N parallel lines which are equally spaced and each data pattern is plotted as a polyline with vertices on the parallel axes. The parallel coordinates visualisation technique was invented by Maurice d'Ocagne [43] and popularised by Alfred Inselberg [94]. Recently, Franken proposed parallel coordinates as an information visualisation technique to show any interdependencies and trends between parameters of a model (if any) [54]. A similar approach is followed in this section.

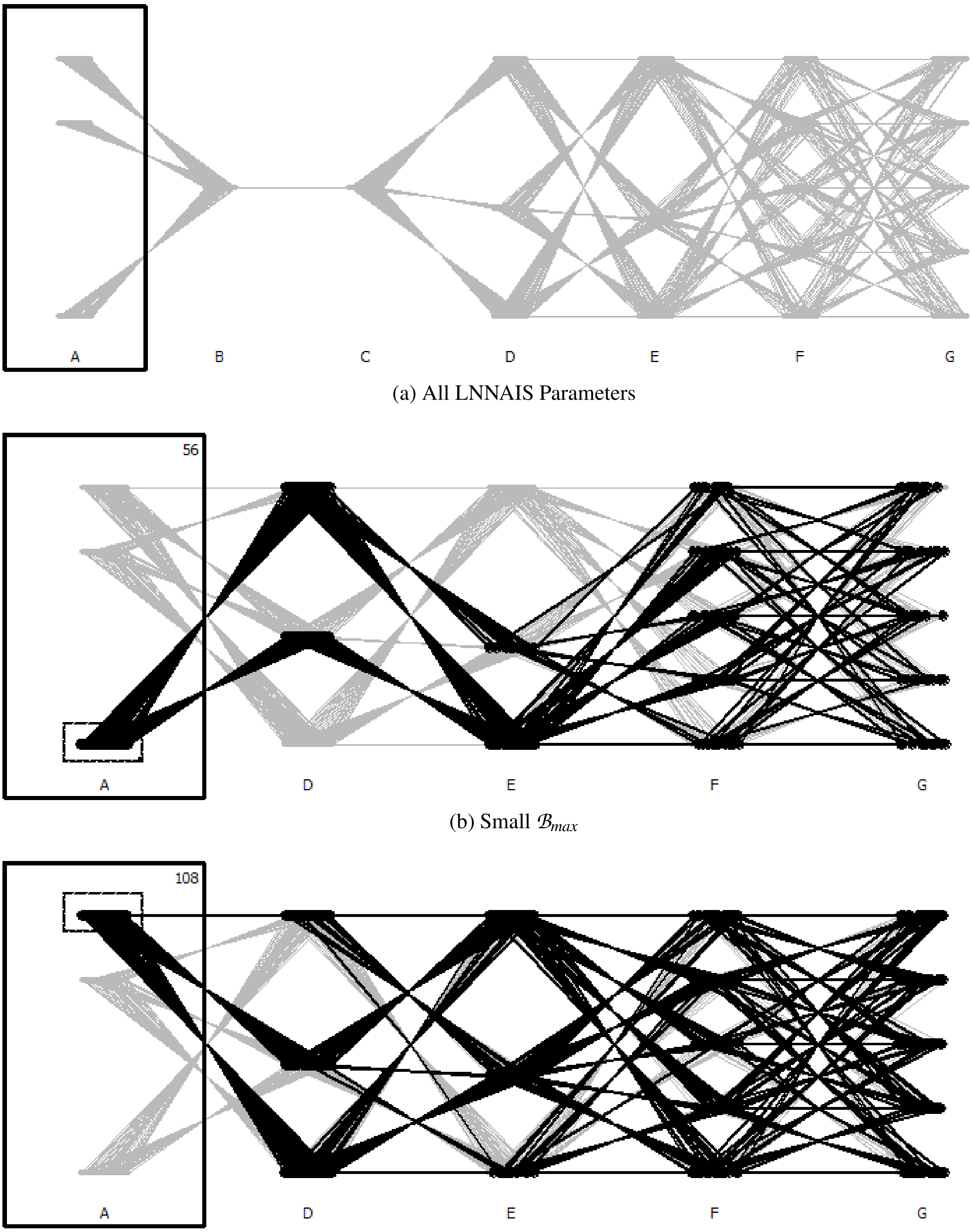
The optimal set of LNNAIS parameter values for each non-stationary data set is plotted with the associated environment parameters which defines the specific non-stationary data set. The LNNAIS parameter values for each data set were found empirically to deliver the best performance. The parallel coordinates plot of the optimal set of LNNAIS parameter values will illustrate the effect of changes in the environment on the optimal set of parameter values of LNNAIS. All parallel coordinates plots in this section consist of axes which are represented by letters A to G. These letters map to the following parameters: $A \mapsto \mathcal{B}_{max}$, $B \mapsto \rho$, $C \mapsto \varepsilon_{clone}$, $D \mapsto N$ (number of dimensions), $E \mapsto |C|$ (cluster size), $F \mapsto \tilde{f}$ (frequency of change) and $G \mapsto \tilde{s}$ (severity of change). The lowest value of each axis in the parallel coordinates plot is at the bottom of the axis. Furthermore, three dimensional plots of the environment parameters versus the clustering quality of LNNAIS illustrate the effect of changes in different environments on the clustering performance of LNNAIS.

7.4.1 Pattern Migration

Figure 7.1(a) illustrates the parallel coordinates plot for pattern migration environments. Figure 7.1(a) shows that there is no change in axes B or C for all values of D to G. This means that changes in the pattern migration environment have no effect on the optimal values of ρ and ε_{clone}. These axes are removed in figures 7.1(b) and 7.1(c) to focus more on the effect of changes in the pattern migration environment on parameter $\mathcal{B}_{max}$. The polylines for small values of $\mathcal{B}_{max}$ in figure 7.1(b) and larger values of $\mathcal{B}_{max}$ in figure 7.1(c) are highlighted. The highlighted polylines show that LNNAIS utilises small population sizes for high dimensional environments with small cluster sizes (illustrated by axes D and E in figure 7.1(b)) and larger population sizes at different dimensions for large cluster sizes (illustrated by axes D and E in figure 7.1(c)). Note that there is no effect on $\mathcal{B}_{max}$ with different frequencies or severities of change. Table 7.1 shows that, in general, the clustering quality of LNNAIS is the lowest at high frequencies and high severities of change in pattern migration environments at different dimensions and cluster sizes. The clustering quality of LNNAIS improves with an increase in the cluster size at different dimensions. Increasing the number of dimensions lowers the clustering quality of LNNAIS at different cluster sizes.

7.4.2 Cluster Migration

Figure 7.2(a) illustrates the parallel coordinates plot for cluster migration environments. Figures 7.2(b) and 7.2(c) respectively illustrates a similar trend in $\mathcal{B}_{max}$ as for pattern migration environments. The highlighted polylines show that LNNAIS utilises small population sizes for high dimensional cluster migration environments with small cluster sizes (illustrated by axes D and E in figure 7.2(b)) and larger population sizes at different dimensions for large cluster sizes (illustrated by axes D and E in figure 7.2(c)). Note that there is also no effect on $\mathcal{B}_{max}$ with different frequencies or severities of change. Table 7.2 shows similar trends on the clustering performance of LNNAIS for cluster migration environments as for pattern migration environments (as shown in table 7.1). In general, the clustering quality of LNNAIS in cluster migration environments is also the lowest at high frequencies and high severities of change at different dimensions and cluster sizes. The clustering quality of LNNAIS improves with an increase in the cluster size at different dimensions and an increase in the number of dimensions lowers the clustering quality of LNNAIS at different cluster sizes.



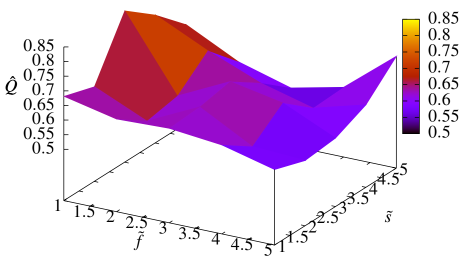
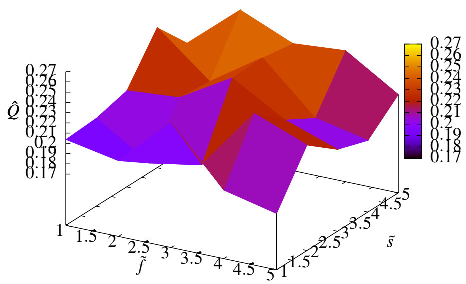
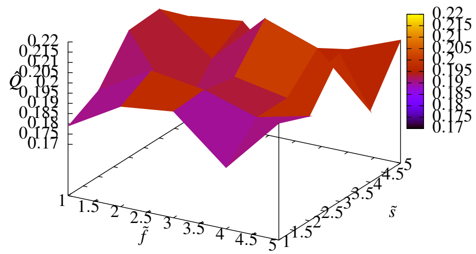
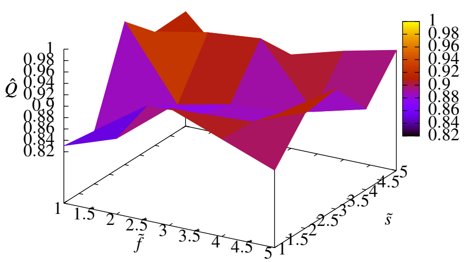
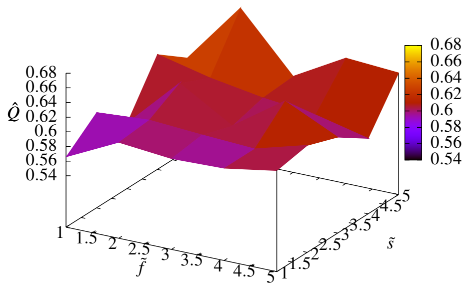
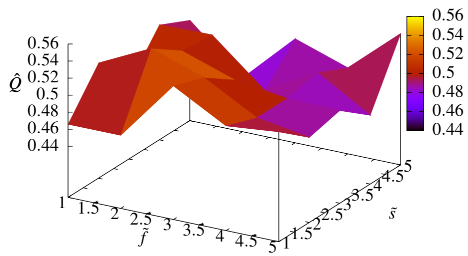
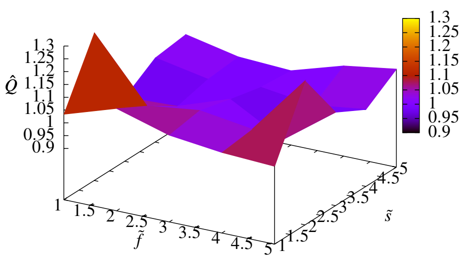
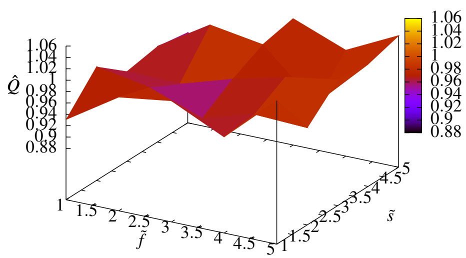
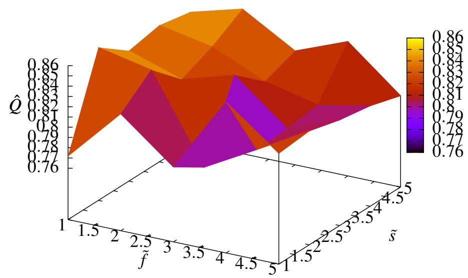
(a) All LNNAIS Parameters

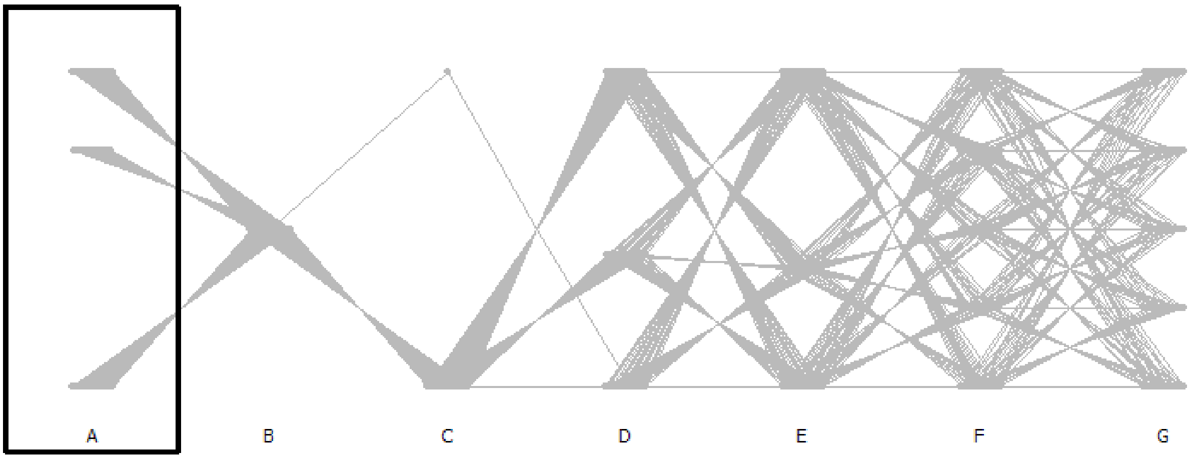
(b) Small $\mathcal{B}_{max}$

(c) Large $\mathcal{B}_{max}$

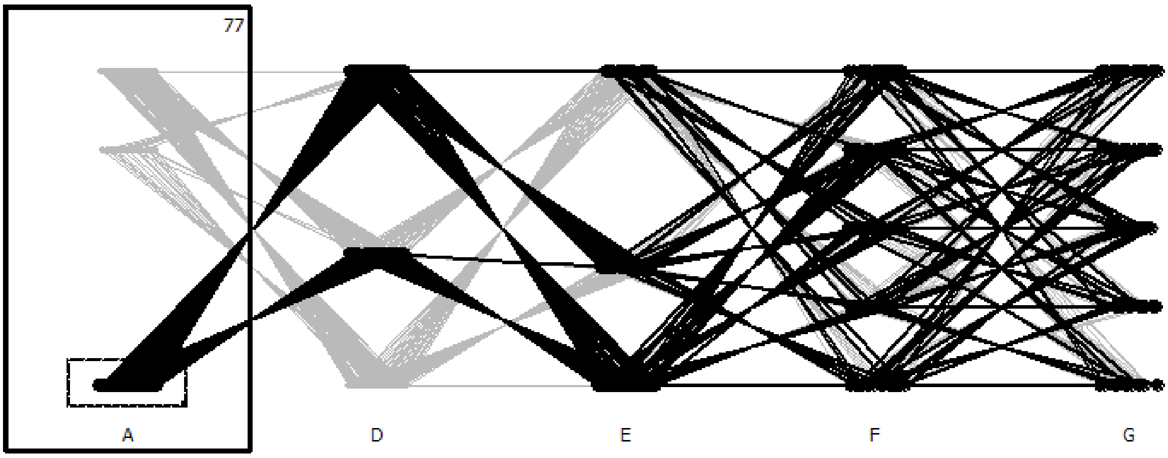
Figure 7.1 Parallel Coordinates of LNNAIS Parameters for Pattern Migration

Table 7.1 Effect of Pattern Migration on Clustering Performance of LNNAIS

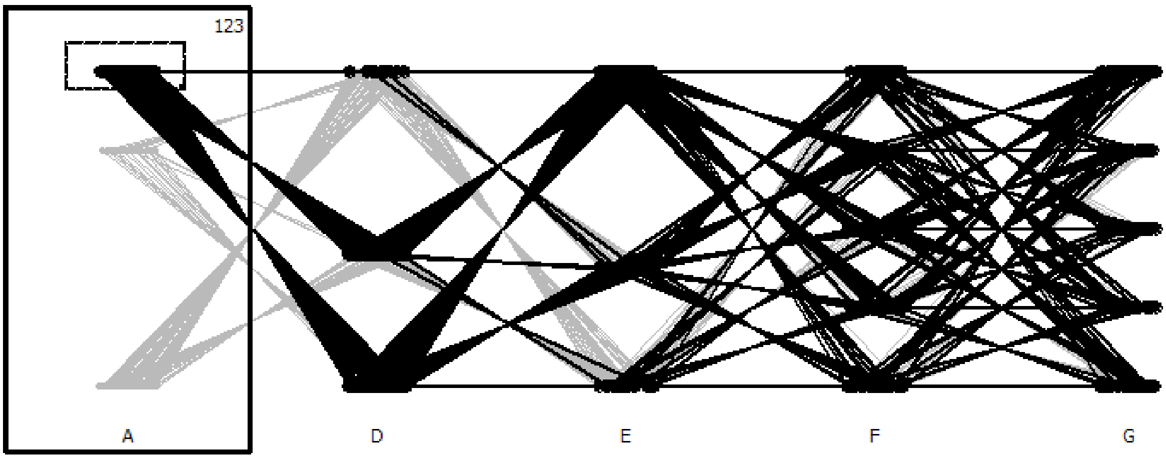
	$\lvert C\rvert = 10$	$\lvert C\rvert = 25$	$\lvert C\rvert = 50$
$N = 3$			
$N = 8$			
$N = 15$			



(a) All LNNAIS Parameters



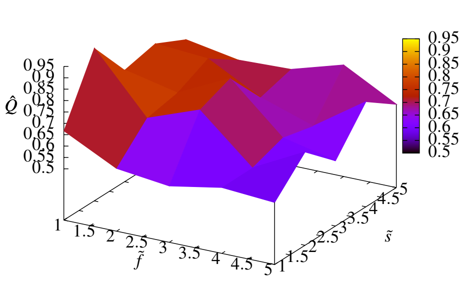
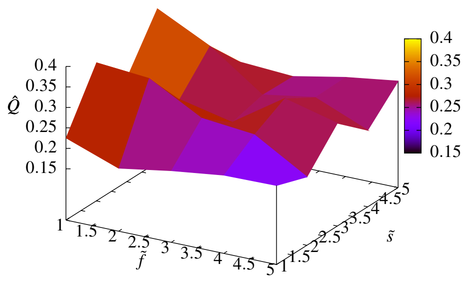
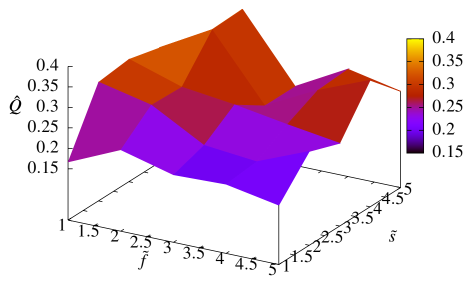
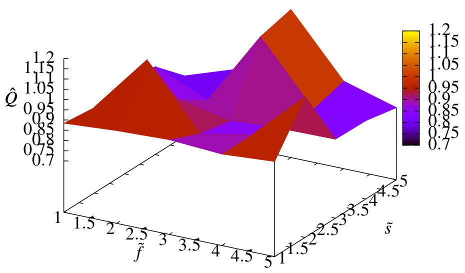
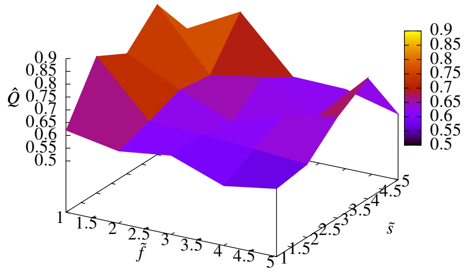
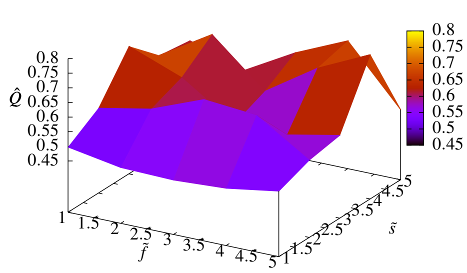
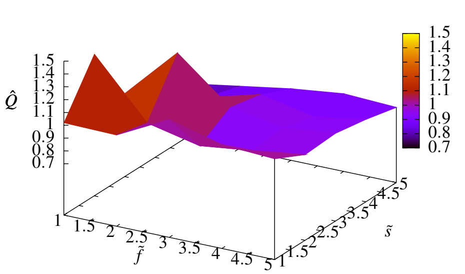
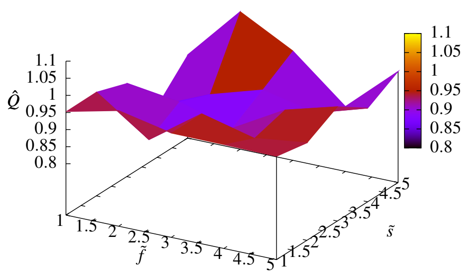
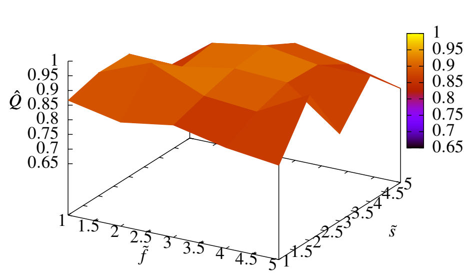
(b) Small $\mathcal{B}_{max}$

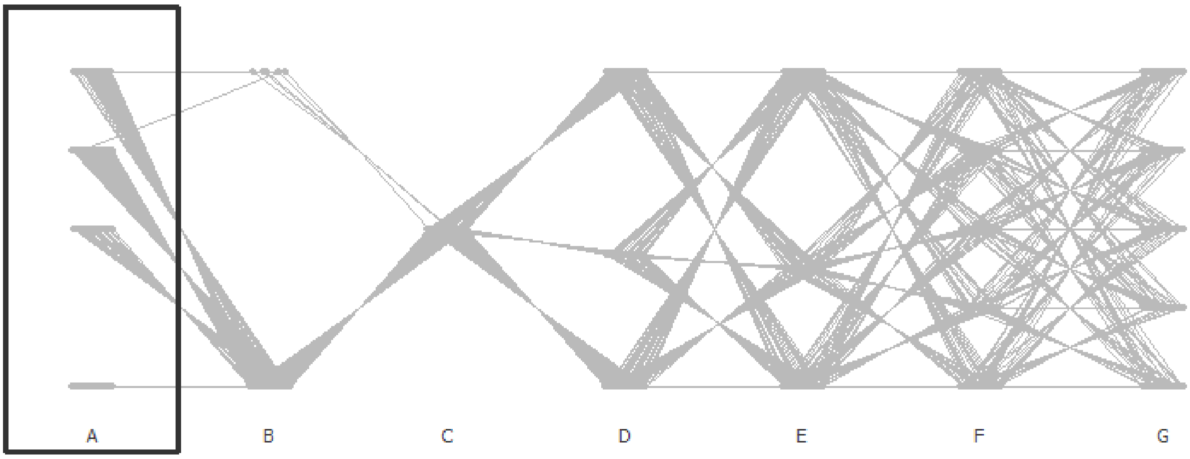


(c) Large $\mathcal{B}_{max}$

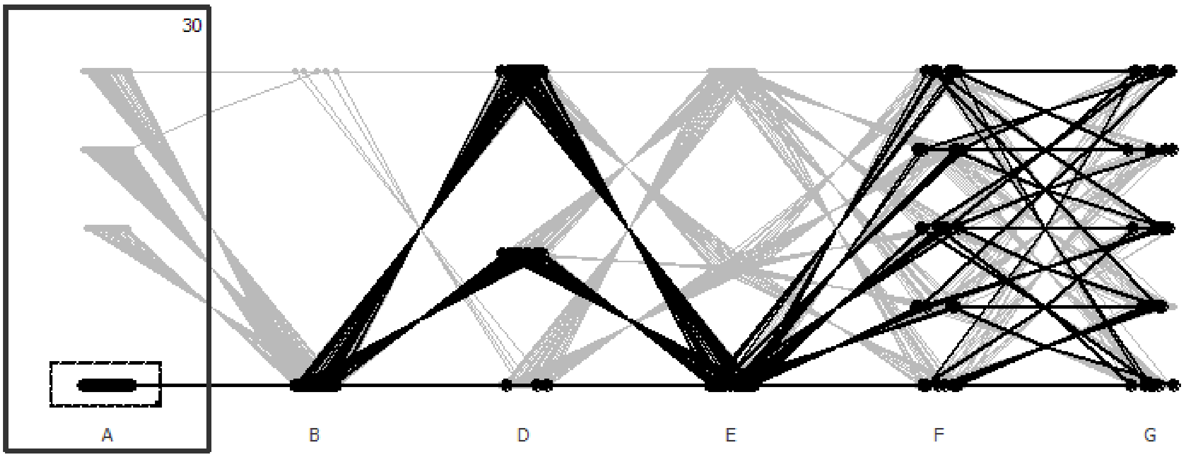
Figure 7.2 Parallel Coordinates of LNNAIS Parameters for Cluster Migration

Table 7.2 Effect of Cluster Migration on Clustering Performance of LNNAIS

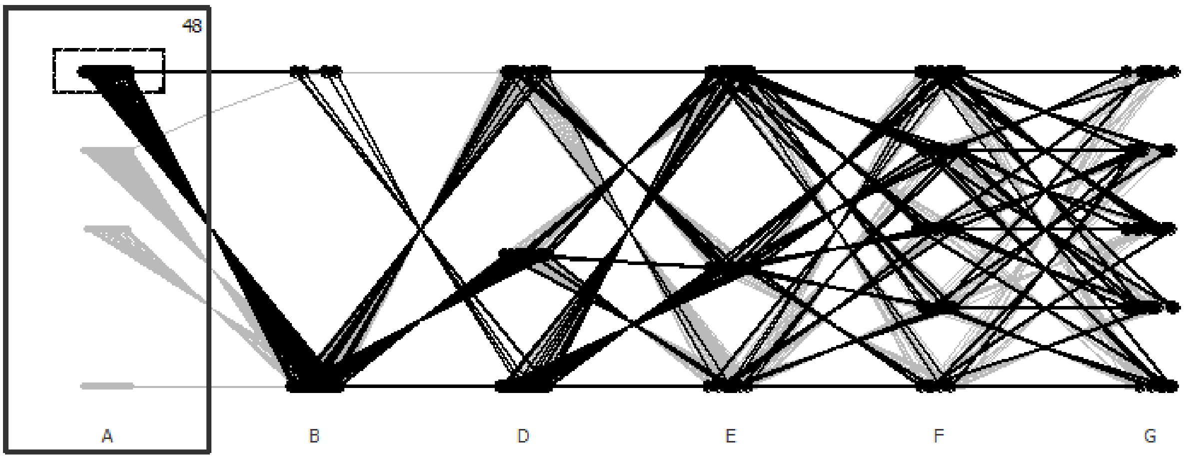
	$\|C\| = 10$	$\|C\| = 25$	$\|C\| = 50$
$N = 3$			
$N = 8$			
$N = 15$			



(a) All LNNAIS Parameters



(b) Small $\mathcal{B}_{max}$



(c) Large $\mathcal{B}_{max}$

Figure 7.3 Parallel Coordinates of LNNAIS Parameters for Centroid Migration

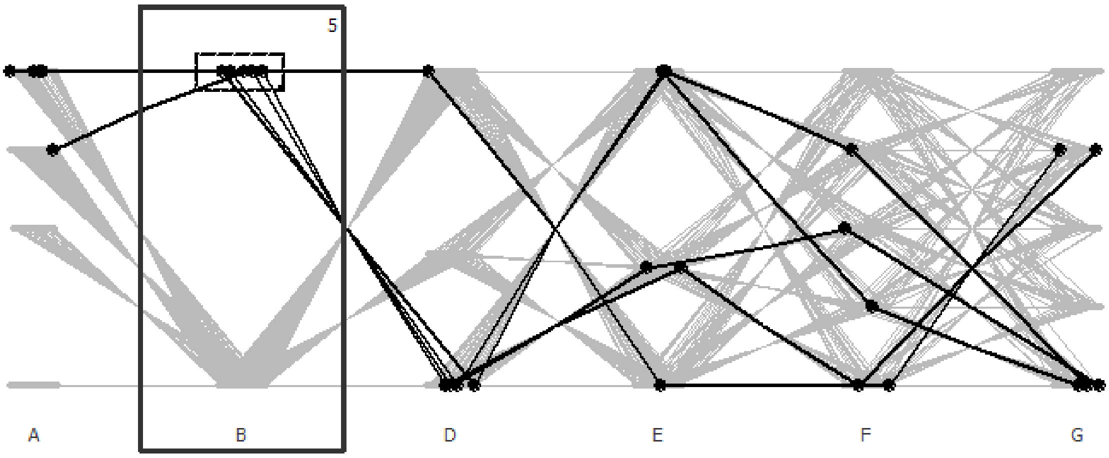
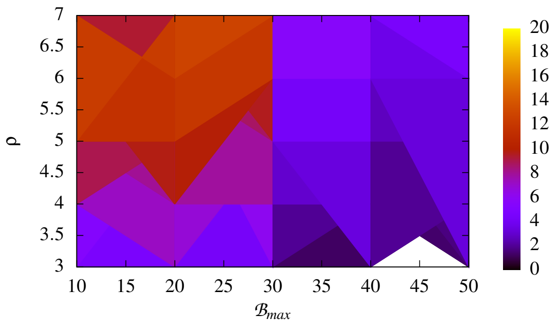


Figure 7.4 Parallel Coordinates of ρ for Centroid Migration

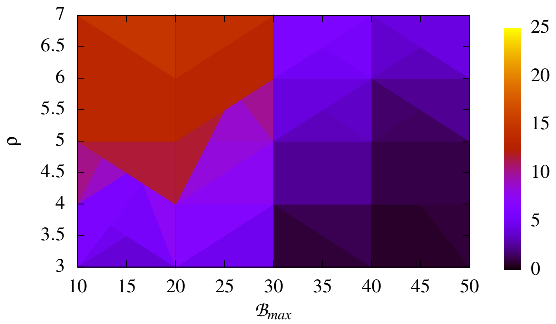
7.4.3 Centroid Migration

Figure 7.3(a) illustrates the parallel coordinates plot for centroid migration environments. Similar to the parallel coordinates plots for pattern and cluster migration environments, there is no effect on the value of ε_{clone} with changes in the centroid migration environment. However, there is an effect on the value of ρ (where the minority of polylines have $\rho = 4$): Figure 7.4 highlights the polylines for $\rho = 4$. The majority of these are for $N = 3$ and are investigated next. Figure 7.5 illustrates the heat maps of the clustering performance of LNNAIS at different values of $\mathcal{B}_{max}$ and ρ for the centroid migration environments which are highlighted in figure 7.4. These heat maps show that there is no distinct difference in the clustering quality of LNNAIS with $\mathcal{B}_{max} = 50$ and $\rho \leq 5$. Therefore these polylines can be seen as outliers to the norm of $\rho = 3$.

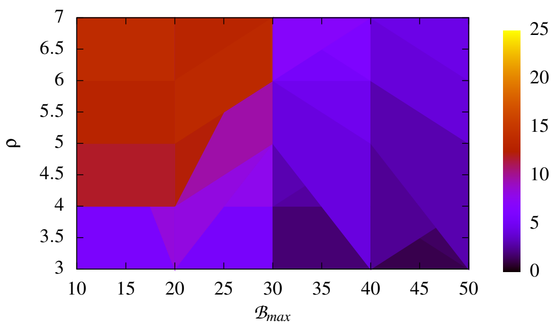
Figures 7.3(b) and 7.3(c) respectively illustrate a similar trend in $\mathcal{B}_{max}$ as for pattern and cluster migration environments. The highlighted polylines show that LNNAIS utilises small population sizes with small cluster sizes (illustrated by axis E in figure 7.3(b)) and larger population sizes for large cluster sizes (illustrated by axis E in figure 7.3(c)). Different to the pattern and cluster migration environments, LNNAIS utilises small and large population sizes at different dimensions. Again note that there is also no effect on $\mathcal{B}_{max}$ with different frequencies or severities of change. Similar trends on the clustering performance of LNNAIS for pattern and cluster migration environments are shown in table 7.3 for centroid migration environments. The clustering quality of LNNAIS in centroid migration environments is the lowest at high frequencies and high severities of change at different dimensions and cluster sizes, the clustering quality of LNNAIS improves with an increase in the cluster size at different dimensions, and an increase in the number of di-



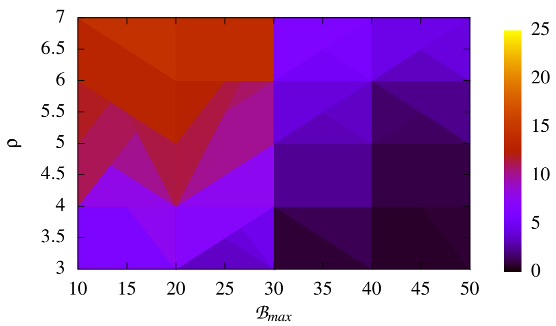
(a) $|C| = 25$, $\tilde{f} = 1$, $\tilde{s} = 4$



(b) $|C| = 50$, $\tilde{f} = 2$, $\tilde{s} = 1$



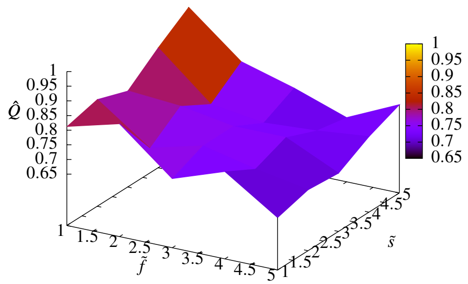
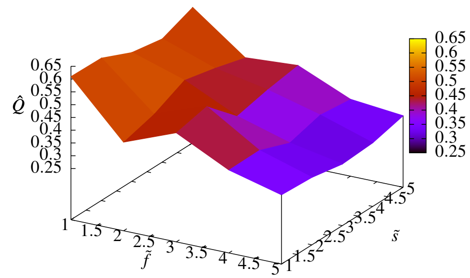
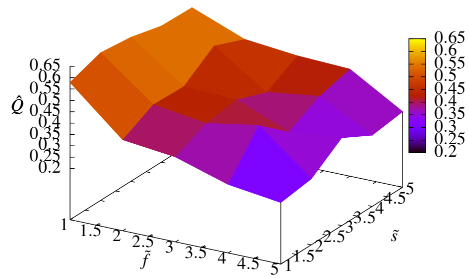
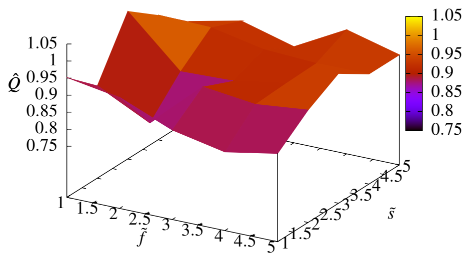
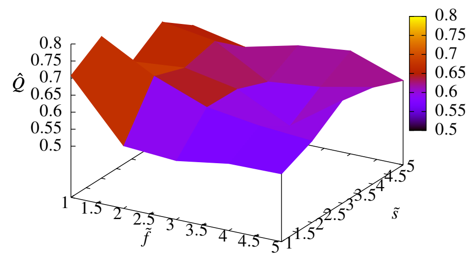
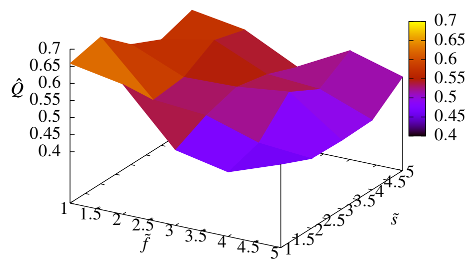
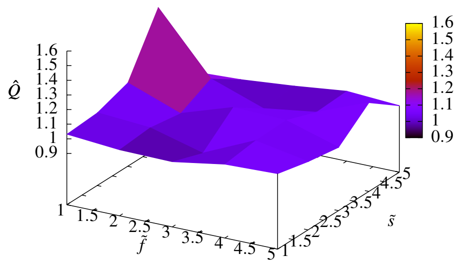
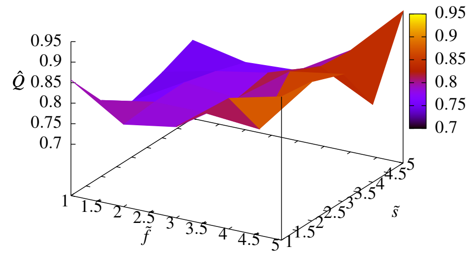
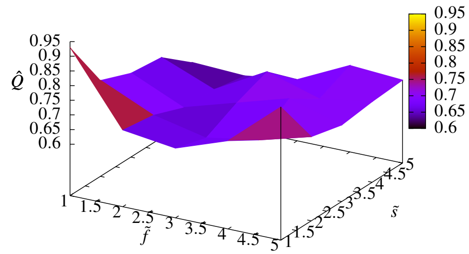
(c) $|C| = 25$, $\tilde{f} = 3$, $\tilde{s} = 1$



(d) $|C| = 50$, $\tilde{f} = 4$, $\tilde{s} = 1$

Figure 7.5 Heat Maps of LNNAIS Parameters for Centroid Migration ($N = 3$)

Table 7.3 Effect of Centroid Migration on Clustering Performance of LNNAIS

	$\|C\| = 10$	$\|C\| = 25$	$\|C\| = 50$
$N = 3$			
$N = 8$			
$N = 15$			

mensions lowers the clustering quality of LNNAIS at different cluster sizes. Note that different to the pattern and cluster migration environments, at high dimensions in the centroid migration environments, the function of frequency versus severity versus clustering quality tends to flatten at different cluster sizes. This shows that the frequency and severity of change in high dimensional centroid migration environments have a smaller effect on the clustering performance of LNNAIS.

7.5 Experimental Results

This section compares the clustering performance of LNNAIS with the clustering performance of LNN_{SDOT}, SMAIN and DWB for each migration type.

The clustering quality of a model for a specific run can be measured using $\bar{Q}_i$ if the data is non-stationary (as defined in equation (7.1)). This section investigates whether there is a difference between the mean clustering quality, $\bar{Q}$, of two models for a specific non-stationary data set or not. The hypothesis can therefore be defined as

- *Null* hypothesis, H_0: There is no difference in $\bar{Q}$.
- *Alternative* hypothesis, H_1: There is a difference in $\bar{Q}$.

A non-parametric Mann-Whitney U test with a 0.95 confidence interval ($\alpha = 0.05$) was used to test the above hypothesis. Furthermore, the clustering quality (Q_{ratio}, J_{intra}, J_{inter} and K) is plotted against time to quantify the quality of partitioning of the non-stationary environment by each clustering algorithm over time.

Due to the prohibitively large number of generated non-stationary data sets, a representative configuration of environment parameter values was selected to compare the clustering performance of the different models. The environment parameters were set to $N = 8$, $|C| = 25$, $\tilde{f} = 3$ and $\tilde{s} = 3$. The value of each parameter in the selected configuration is then changed while the remaining parameter values are kept constant. Therefore each migration type has five data sets representing different severities of change (with the remaining environment parameters kept constant), five data sets representing different frequencies of change, three data sets representing different dimensions and three data sets representing different cluster sizes. This gives a total of sixteen different non-stationary data sets for each migration type. The results for each of these

Table 7.4 LNNAIS Parameter Values - Pattern Migration

Environment parameters	$\mathcal{B}_{max}$	ρ	ε_{clone}
$N = 3$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 3$	50	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 3$	50	3	5
$N = 15$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 3$	50	3	5
$N = 8$; $\|C\| = 10$; $\tilde{f} = 3$; $\tilde{s} = 3$	10	3	5
$N = 8$; $\|C\| = 50$; $\tilde{f} = 3$; $\tilde{s} = 3$	40	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 1$; $\tilde{s} = 3$	50	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 2$; $\tilde{s} = 3$	40	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 4$; $\tilde{s} = 3$	50	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 5$; $\tilde{s} = 3$	40	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 1$	50	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 2$	50	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 4$	50	3	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 5$	50	3	5

migration types are discussed next. The parameter values used by each model for each of the migration types are summarised in tables 7.4 - 7.12 for LNNAIS, DWB and SMAIN, respectively.

7.5.1 Pattern Migration

Figure 7.6 illustrates the quality of partitioning by the different models over time for pattern migration. Note the increase in the ALC population size for the SMAIN model with every change in the data (figure 7.6(d) at $t = 30$, $t = 60$ and $t = 90$). Figure 7.6(a) shows that LNNAIS initially finds clusters with a lower quality (high Q_{ratio} value) when compared to the other models. As time progresses, the quality of the clusters found by LNNAIS improves (as illustrated in figures 7.6(b), 7.6(c) and 7.6(e) the number of clusters found increases, becomes more compact and separated). LNNAIS delivers clusters of a higher quality than DWB and LNN$_{SDOT}$ for all pattern migration environments. Figure 7.6(f) illustrates the average number of clusters and the number of clusters with the highest frequency which was dynamically determined by LNN$_{SDOT}$. The number of clusters detected with the highest frequency was $K = 3$ at every point in time with an average number of clusters between $K = 3$ and $K = 3.6$. Although LNN$_{SDOT}$ was unable to detect the correct number of clusters ($K = 8$), figure 7.6(f) shows that there was no change in the number of clusters over time, which is expected for pattern migration where data patterns randomly migrate between a static number of clusters. Section 7.5.3 discusses the argument for

Table 7.5 LNNAIS Parameter Values - Cluster Migration

Environment parameters	$\mathcal{B}_{max}$	ρ	ε_{clone}
$N=3$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	50	3	5
$N=15$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	10	3	5
$N=8$; $\|C\|=10$; $\tilde{f}=3$; $\tilde{s}=3$	10	3	5
$N=8$; $\|C\|=50$; $\tilde{f}=3$; $\tilde{s}=3$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=1$; $\tilde{s}=3$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=2$; $\tilde{s}=3$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=4$; $\tilde{s}=3$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=5$; $\tilde{s}=3$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=1$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=2$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=4$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=5$	50	3	5

Table 7.6 LNNAIS Parameter Values - Centroid Migration

Environment parameters	$\mathcal{B}_{max}$	ρ	ε_{clone}
$N=3$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	50	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	50	3	5
$N=15$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	30	3	5
$N=8$; $\|C\|=10$; $\tilde{f}=3$; $\tilde{s}=3$	10	3	5
$N=8$; $\|C\|=50$; $\tilde{f}=3$; $\tilde{s}=3$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=1$; $\tilde{s}=3$	30	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=2$; $\tilde{s}=3$	30	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=4$; $\tilde{s}=3$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=5$; $\tilde{s}=3$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=1$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=2$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=4$	40	3	5
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=5$	40	3	5

Table 7.7 DWB Parameter Values - Pattern Migration

Environment parameters	$\mathcal{B}_{max}$	ϕ_{init}	m_{min}	A	a_{min}	a_{max}	k_{clone}	ς	τ	τ_α	τ_β	$k_{compress}$
$N=3$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	84	0.705	0.114	55	6	6	6	0.93	8	8	8	6
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	34	0.184	0.634	91	23	41	12	0.297	9	3	6	6
$N=15$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	29	0.262	0.43	73	74	74	9	0.571	10	4	6	5
$N=8$; $\lvert C\rvert=10$; $\tilde{f}=3$; $\tilde{s}=3$	43	0.213	0.213	38	63	63	10	0.888	7	2	7	4
$N=8$; $\lvert C\rvert=50$; $\tilde{f}=3$; $\tilde{s}=3$	44	0.648	0.395	12	24	24	7	0.536	7	1	5	3
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=1$; $\tilde{s}=3$	32	0.986	0.733	74	37	37	2	0.423	8	9	6	1
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=2$; $\tilde{s}=3$	40	0.149	0.543	85	13	13	4	0.459	4	3	3	6
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=4$; $\tilde{s}=3$	44	0.648	0.395	12	24	24	7	0.536	7	1	5	3
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=5$; $\tilde{s}=3$	37	0.402	0.234	95	17	71	4	0.655	5	8	9	1
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=1$	34	0.184	0.634	91	23	41	12	0.297	9	3	6	6
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=2$	37	0.402	0.234	95	17	71	4	0.655	5	8	9	1
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=4$	41	0.508	0.873	15	15	15	13	0.62	5	7	6	6
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=5$	37	0.606	0.944	44	7	7	2	0.944	4	5	6	4

Table 7.8 DWB Parameter Values - Cluster Migration

Environment parameters	$\mathcal{B}_{max}$	ϕ_{init}	m_{min}	A	a_{min}	a_{max}	k_{clone}	ς	τ	τ_α	τ_β	$k_{compress}$
$N = 3$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 3$	95	0.74	0.346	82	30	34	13	0.318	6	9	6	7
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 3$	27	0.81	0.501	93	47	54	5	0.895	8	8	6	2
$N = 15$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 3$	31	0.325	0.775	75	26	75	12	0.775	8	8	3	3
$N = 8$; $\|C\| = 10$; $\tilde{f} = 3$; $\tilde{s} = 3$	35	0.677	0.93	58	77	95	1	0.677	10	2	4	6
$N = 8$; $\|C\| = 50$; $\tilde{f} = 3$; $\tilde{s} = 3$	31	0.325	0.775	75	26	75	12	0.775	8	8	3	3
$N = 8$; $\|C\| = 25$; $\tilde{f} = 1$; $\tilde{s} = 3$	29	0.262	0.43	73	74	74	9	0.571	10	4	6	5
$N = 8$; $\|C\| = 25$; $\tilde{f} = 2$; $\tilde{s} = 3$	31	0.325	0.775	75	26	75	12	0.775	8	8	3	3
$N = 8$; $\|C\| = 25$; $\tilde{f} = 4$; $\tilde{s} = 3$	32	0.986	0.733	74	37	37	2	0.423	8	9	6	1
$N = 8$; $\|C\| = 25$; $\tilde{f} = 5$; $\tilde{s} = 3$	27	0.81	0.501	93	47	54	5	0.895	8	8	6	2
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 1$	44	0.648	0.395	12	24	24	7	0.536	7	1	5	3
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 2$	36	0.445	0.304	10	33	33	3	0.81	9	9	2	6
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 4$	36	0.445	0.304	10	33	33	3	0.81	9	9	2	6
$N = 8$; $\|C\| = 25$; $\tilde{f} = 3$; $\tilde{s} = 5$	39	0.923	0.388	56	17	17	14	0.107	5	7	2	4

Table 7.9 DWB Parameter Values - Centroid Migration

Environment parameters	$\mathcal{B}_{max}$	ϕ_{init}	m_{min}	A	a_{min}	a_{max}	k_{clone}	ς	τ	τ_α	τ_β	$k_{compress}$
$N=3$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	95	0.74	0.346	82	30	34	13	0.318	6	9	6	7
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	53	0.838	0.923	3	31	88	9	0.866	3	6	2	7
$N=15$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	68	0.747	0.297	4	10	29	14	0.409	8	6	7	5
$N=8$; $\lvert C\rvert=10$; $\tilde{f}=3$; $\tilde{s}=3$	53	0.17	0.311	52	27	27	11	0.733	1	4	5	7
$N=8$; $\lvert C\rvert=50$; $\tilde{f}=3$; $\tilde{s}=3$	53	0.838	0.923	3	31	88	9	0.866	3	6	2	7
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=1$; $\tilde{s}=3$	95	0.74	0.346	82	30	34	13	0.318	6	9	6	7
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=2$; $\tilde{s}=3$	53	0.838	0.923	3	31	88	9	0.866	3	6	2	7
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=4$; $\tilde{s}=3$	40	0.149	0.543	85	13	13	4	0.459	4	3	3	6
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=5$; $\tilde{s}=3$	68	0.747	0.297	4	10	29	14	0.409	8	6	7	5
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=1$	46	0.768	0.712	79	68	68	8	0.515	7	2	1	7
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=2$	69	0.543	0.937	54	14	93	12	0.29	9	4	8	3
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=4$	69	0.543	0.937	54	14	93	12	0.29	9	4	8	3
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=5$	55	0.311	0.958	98	12	12	5	0.198	6	7	3	7

Table 7.10 SMAIN Parameter Values - Pattern Migration

Environment parameters	$\mathcal{B}_{init}$	R_γ	R_Λ	NAT	R_k	R_{max}	R_{init}
$N=3$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.042	0.824	22	1.445	0.852	819	31
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.057	0.73	3	3.238	0.539	269	62
$N=15$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.122	0.828	29	4.872	0.531	875	47
$N=8$; $\|C\|=10$; $\tilde{f}=3$; $\tilde{s}=3$	0.198	0.797	10	2.566	0.719	925	29
$N=8$; $\|C\|=50$; $\tilde{f}=3$; $\tilde{s}=3$	0.147	0.793	40	3.751	0.914	969	50
$N=8$; $\|C\|=25$; $\tilde{f}=1$; $\tilde{s}=3$	0.074	0.992	71	3.911	0.734	488	72
$N=8$; $\|C\|=25$; $\tilde{f}=2$; $\tilde{s}=3$	0.098	0.848	5	3.559	0.68	606	98
$N=8$; $\|C\|=25$; $\tilde{f}=4$; $\tilde{s}=3$	0.115	0.969	7	4.231	0.938	550	44
$N=8$; $\|C\|=25$; $\tilde{f}=5$; $\tilde{s}=3$	0.241	0.52	20	3.687	0.711	831	21
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=1$	0.216	0.977	37	3.527	0.078	713	38
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=2$	0.203	0.941	73	3.302	0.492	856	74
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=4$	0.154	0.902	25	3.366	0.57	494	47
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=5$	0.154	0.902	25	3.366	0.57	494	47

Table 7.11 SMAIN Parameter Values - Cluster Migration

Environment parameters	$\mathcal{B}_{init}$	R_γ	R_Λ	NAT	R_k	R_{max}	R_{init}
$N=3$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.226	0.926	2	1.38	0.898	281	76
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.057	0.73	3	3.238	0.539	269	62
$N=15$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.244	0.965	13	4.904	0.195	994	84
$N=8$; $\|C\|=10$; $\tilde{f}=3$; $\tilde{s}=3$	0.053	0.629	11	2.79	0.367	956	54
$N=8$; $\|C\|=50$; $\tilde{f}=3$; $\tilde{s}=3$	0.182	0.703	4	3.847	0.281	675	23
$N=8$; $\|C\|=25$; $\tilde{f}=1$; $\tilde{s}=3$	0.147	0.793	40	3.751	0.914	969	50
$N=8$; $\|C\|=25$; $\tilde{f}=2$; $\tilde{s}=3$	0.216	0.977	37	3.527	0.078	713	38
$N=8$; $\|C\|=25$; $\tilde{f}=4$; $\tilde{s}=3$	0.091	0.957	33	3.174	0.836	331	95
$N=8$; $\|C\|=25$; $\tilde{f}=5$; $\tilde{s}=3$	0.203	0.941	73	3.302	0.492	856	74
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=1$	0.203	0.941	73	3.302	0.492	856	74
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=2$	0.057	0.73	3	3.238	0.539	269	62
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=4$	0.098	0.848	5	3.559	0.68	606	98
$N=8$; $\|C\|=25$; $\tilde{f}=3$; $\tilde{s}=5$	0.147	0.793	40	3.751	0.914	969	50

Table 7.12 SMAIN Parameter Values - Centroid Migration

Environment parameters	$\mathcal{B}_{init}$	R_γ	R_Λ	NAT	R_k	R_{max}	R_{init}
$N=3$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.021	0.945	18	2.245	0.516	363	77
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.036	0.602	12	4.552	0.828	913	13
$N=15$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=3$	0.22	0.812	63	4.488	0.125	300	64
$N=8$; $\lvert C\rvert=10$; $\tilde{f}=3$; $\tilde{s}=3$	0.115	0.969	7	4.231	0.938	550	44
$N=8$; $\lvert C\rvert=50$; $\tilde{f}=3$; $\tilde{s}=3$	0.036	0.602	12	4.552	0.828	913	13
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=1$; $\tilde{s}=3$	0.151	0.582	8	4.199	0.086	531	70
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=2$; $\tilde{s}=3$	0.158	0.723	30	4.584	0.43	806	73
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=4$; $\tilde{s}=3$	0.158	0.723	30	4.584	0.43	806	73
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=5$; $\tilde{s}=3$	0.115	0.969	7	4.231	0.938	550	44
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=1$	0.036	0.602	12	4.552	0.828	913	13
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=2$	0.158	0.723	30	4.584	0.43	806	73
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=4$	0.244	0.965	13	4.904	0.195	994	84
$N=8$; $\lvert C\rvert=25$; $\tilde{f}=3$; $\tilde{s}=5$	0.212	0.641	16	4.359	0.906	375	60

LNN$_{SDOT}$ not being able to detect the correct number of clusters for pattern and cluster migration data sets. Overall, there is no significant change in Q_{ratio}, J_{intra}, J_{inter} and K for all the models ($t > 40$) in pattern migration environments.

The Mann-Whitney U statistical hypothesis test rejects H_0 that the mean clustering quality, $\bar{Q}$, are the same at a 0.05 level of significance between LNNAIS and LNN$_{SDOT}$, SMAIN and DWB for different dimensions (as summarised in table 7.13), different clusters sizes (as summarised in table 7.14), for all frequencies of change (as summarised in table 7.15) and all severities of change (as summarised in table 7.16). There is thus a statistical significant difference in the clustering quality of all the pattern migration data sets between LNNAIS and all the other models.

Table 7.13: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Pattern migration at different dimensions ($|C|$=25; $\tilde{f}$=3; $\tilde{s}$=3)

N	LNN$_{SDOT}$	DWB	SMAIN
3	$z = 3.999$	$z = 6.646$	$z = 6.646$
3	$p < 0.001$	$p < 0.001$	$p < 0.001$
3	Reject H_0	Reject H_0	Reject H_0
8	$z = 6.646$	$z = 6.446$	$z = 6.402$

Continued on next page

N	LNN_{SDOT}	DWB	SMAIN
8	$p < 0.001$	$p < 0.001$	$p < 0.001$
8	Reject H_0	Reject H_0	Reject H_0
15	$z = 6.646$	$z = 5.988$	$z = 6.646$
15	$p < 0.001$	$p < 0.001$	$p < 0.001$
15	Reject H_0	Reject H_0	Reject H_0

Table 7.14: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Pattern migration at different cluster sizes (N=8; $\tilde{s}$=3; $\tilde{f}$=3)

$\lvert C \rvert$	LNN_{SDOT}	DWB	SMAIN
	$z = 6.084$	$z = 6.646$	$z = 6.646$
10	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Reject H_0	Reject H_0	Reject H_0
	$z = 6.646$	$z = 6.446$	$z = 6.402$
25	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Reject H_0	Reject H_0	Reject H_0
	$z = 6.542$	$z = 6.646$	$z = 6.224$
50	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Reject H_0	Reject H_0	Reject H_0

Table 7.15: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Pattern migration at different frequencies of change (N=8; $\lvert C \rvert$=25; $\tilde{s}$=3)

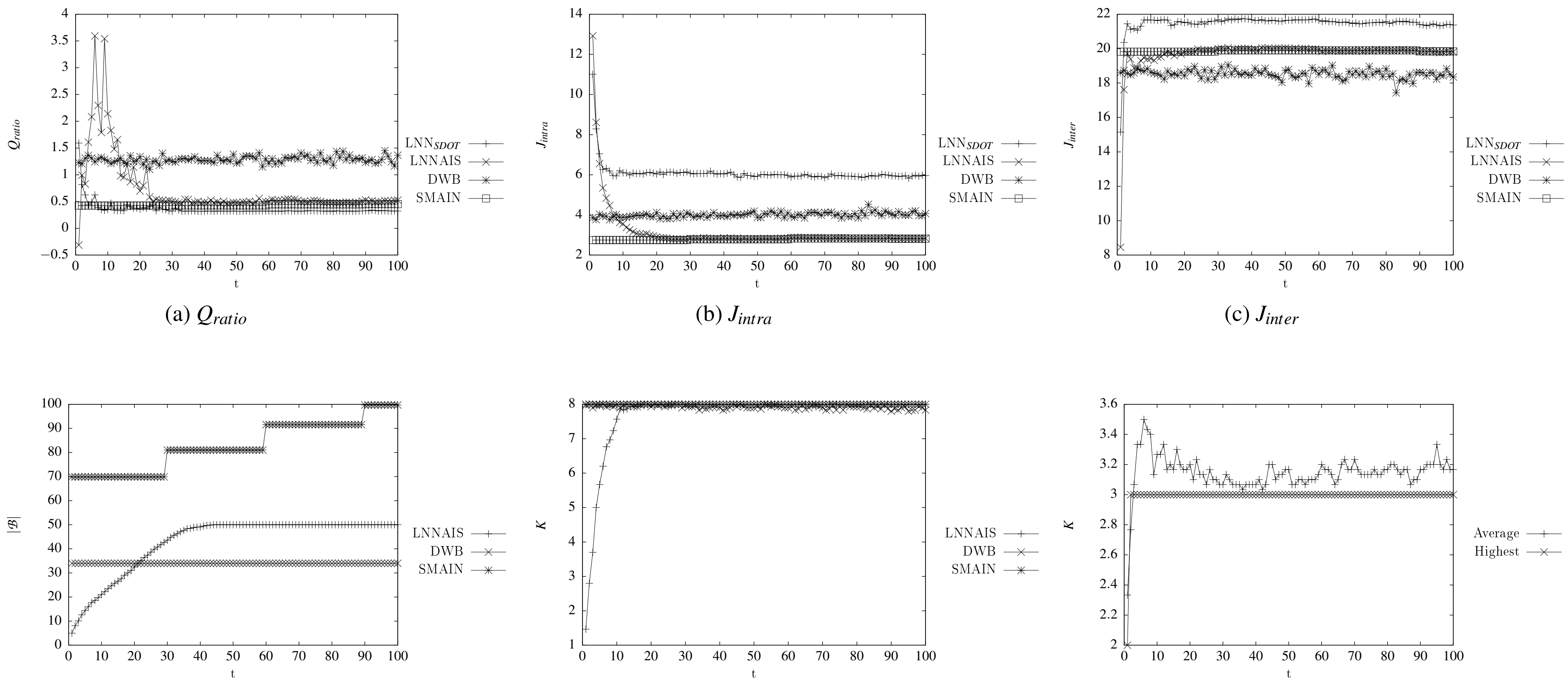
$\tilde{f}$	LNN_{SDOT}	DWB	SMAIN
	$z = 6.557$	$z = 6.646$	$z = 6.572$
1	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Reject H_0	Reject H_0	Reject H_0
	$z = 6.246$	$z = 6.646$	$z = 6.646$

Continued on next page

$\tilde{f}$	LNN_{SDOT}	DWB	SMAIN
2	$p < 0.001$ Reject H_0	$p < 0.001$ Reject H_0	$p < 0.001$ Reject H_0
3	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.446$ $p < 0.001$ Reject H_0	$z = 6.402$ $p < 0.001$ Reject H_0
4	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.439$ $p < 0.001$ Reject H_0	$z = 6.594$ $p < 0.001$ Reject H_0
5	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.557$ $p < 0.001$ Reject H_0

Table 7.16: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Pattern migration at different severities of change (N=8; $|C|$=25; $\tilde{f}$=3)

$\tilde{s}$	LNN_{SDOT}	DWB	SMAIN
1	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.409$ $p < 0.001$ Reject H_0
2	$z = 6.106$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.15$ $p < 0.001$ Reject H_0
3	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.446$ $p < 0.001$ Reject H_0	$z = 6.402$ $p < 0.001$ Reject H_0
4	$z = 6.439$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.616$ $p < 0.001$ Reject H_0
5	$z = 6.402$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.416$ $p < 0.001$ Reject H_0



(a) Q_{ratio} (b) J_{intra} (c) J_{inter}

(d) $|\mathcal{B}|$ (e) K (f) K (LNN$_{SDOT}$)

Figure 7.6 Pattern Migration ($N = 8, |C| = 25, \tilde{f} = 3, \tilde{s} = 3$): Quantifying each model's partitioning quality with regards to Q_{ratio}, $|\mathcal{B}|$, J_{intra}, J_{inter} and K

Although SMAIN tends to find clusters with a higher quality than LNNAIS for different dimensions (see table 7.17), cluster sizes (see table 7.18), frequencies of change (see table 7.19) and severities of change (see table 7.20), SMAIN utilises a larger ALC population size than LNNAIS for all pattern migration data sets. The larger ALC population size which is utilised by SMAIN is an indication of overfitting of the data which results in clusters of higher quality for pattern migration environments. This drawback can have a major impact in the scalability of the SMAIN model where the number of clusters changes (cluster migration environments) and where the centroids of clusters are non-stationary (centroid migration environments).

Table 7.17: Descriptive Statistics: Pattern migration at different dimensions ($|C|$=25; $\tilde{f}$=3; $\tilde{s}$=3)

N		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	K	5.28	8	7.96	8
		(±)2.06	(±)0	(±)0.02	(±)0
	J_{intra}	5.318	1.826	3.245	1.667
		(±)2.787	(±)0.212	(±)0.08	(±)0.013
3	J_{inter}	21.349	21.616	20.095	21.643
		(±)0.782	(±)0.137	(±)0.117	(±)0.047
	$\hat{Q}$	0.368	0.248	1.578	0.151
		(±)0.135	(±)0.168	(±)0.083	(±)0.002
	$\|\mathcal{B}\|$	49.3	49.3	84	84.51
		(±)0.45	(±)0.45	(±)0	(±)2.3
	K	3.14	7.99	7.92	8
		(±)0.32	(±)0.01	(±)0.04	(±)0
	J_{intra}	5.997	2.863	4.004	2.793
		(±)0.348	(±)0.07	(±)0.117	(±)0.014
8	J_{inter}	21.557	19.887	18.53	19.869
		(±)0.219	(±)0.068	(±)0.32	(±)0.084
	$\hat{Q}$	0.335	0.586	1.286	0.436
		(±)0.029	(±)0.183	(±)0.094	(±)0.012
	$\|\mathcal{B}\|$	45.97	45.97	34	83
		(±)1.26	(±)1.26	(±)0	(±)2.89

Continued on next page

N		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	K	4.75	7.97	7.92	7.84
		(±)0.74	(±)0.05	(±)0.06	(±)0.23
	J_{intra}	5.699	4.059	5.347	4.266
		(±)0.803	(±)0.064	(±)0.19	(±)0.068
15	J_{inter}	22.169	20.565	19.685	20.607
		(±)0.464	(±)0.106	(±)0.312	(±)0.224
	$\hat{Q}$	0.425	0.973	1.225	0.663
		(±)0.102	(±)0.117	(±)0.059	(±)0.015
	$\|\mathcal{B}\|$	43.79	43.79	29	84.18
		(±)1.67	(±)1.67	(±)0	(±)3.37

Table 7.18: Descriptive Statistics: Pattern migration at different cluster sizes (N=8; $\tilde{s}$=3; $\tilde{f}$=3)

$\|C\|$		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	K	3.09	3.66	7.92	8
		(±)0.09	(±)0.07	(±)0.04	(±)0
	J_{intra}	6.508	5.99	3.619	2.582
		(±)0.157	(±)0.13	(±)0.055	(±)0.005
10	J_{inter}	21.098	19.888	18.606	19.972
		(±)0.125	(±)0.137	(±)0.197	(±)0.04
	$\hat{Q}$	0.497	0.874	1.194	0.408
		(±)0.173	(±)0.096	(±)0.038	(±)0.006
	$\|\mathcal{B}\|$	9.99	9.99	43	71.51
		(±)0.02	(±)0.02	(±)0	(±)1.53
	K	3.14	7.99	7.92	8
		(±)0.32	(±)0.01	(±)0.04	(±)0
	J_{intra}	5.997	2.863	4.004	2.793
		(±)0.348	(±)0.07	(±)0.117	(±)0.014
25	J_{inter}	21.557	19.887	18.53	19.869
		(±)0.219	(±)0.068	(±)0.32	(±)0.084
	$\hat{Q}$	0.335	0.586	1.286	0.436

Continued on next page

$\|C\|$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
		(±)0.029	(±)0.183	(±)0.094	(±)0.012
	$\|\mathcal{B}\|$	45.97	45.97	34	83
		(±)1.26	(±)1.26	(±)0	(±)2.89
50	K	3.11	8	7.95	8
		(±)0.31	(±)0	(±)0.03	(±)0
	J_{intra}	6.018	2.84	3.915	2.819
		(±)0.307	(±)0.019	(±)0.105	(±)0.017
	J_{inter}	21.683	19.965	18.546	20.107
		(±)0.267	(±)0.035	(±)0.334	(±)0.107
	$\hat{Q}$	0.326	0.499	1.296	0.439
		(±)0.044	(±)0.043	(±)0.06	(±)0.015
	$\|\mathcal{B}\|$	39.7	39.7	44	72.02
		(±)0.28	(±)0.28	(±)0	(±)3.92

Table 7.19: Descriptive Statistics: Pattern migration at different frequencies of change (N=8; $|C|$=25; $\tilde{s}$=3)

$\tilde{f}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
1	K	3.21	7.99	7.94	8
		(±)0.59	(±)0.02	(±)0.03	(±)0
	J_{intra}	6.021	2.891	4.077	2.891
		(±)0.519	(±)0.034	(±)0.077	(±)0.019
	J_{inter}	21.934	20.158	18.772	20.226
		(±)0.342	(±)0.051	(±)0.199	(±)0.093
	$\hat{Q}$	0.335	0.559	1.287	0.435
		(±)0.049	(±)0.089	(±)0.054	(±)0.015
	$\|\mathcal{B}\|$	46.07	46.07	32	61.01
		(±)1.36	(±)1.36	(±)0	(±)2.78
	K	3.09	8	7.94	8
		(±)0.29	(±)0.01	(±)0.03	(±)0
	J_{intra}	6.118	2.861	3.87	2.825
		(±)0.313	(±)0.047	(±)0.071	(±)0.016

Continued on next page

$\tilde{f}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
2	J_{inter}	21.967	19.965	18.67	20.253
		(±)0.255	(±)0.116	(±)0.225	(±)0.105
	$\hat{Q}$	0.335	0.629	1.287	0.409
		(±)0.082	(±)0.111	(±)0.054	(±)0.009
	$\|\mathcal{B}\|$	38.4	38.4	40	64.06
		(±)0.66	(±)0.66	(±)0	(±)2.77
	K	3.14	7.99	7.92	8
		(±)0.32	(±)0.01	(±)0.04	(±)0
	J_{intra}	5.997	2.863	4.004	2.793
		(±)0.348	(±)0.07	(±)0.117	(±)0.014
3	J_{inter}	21.557	19.887	18.53	19.869
		(±)0.219	(±)0.068	(±)0.32	(±)0.084
	$\hat{Q}$	0.335	0.586	1.286	0.436
		(±)0.029	(±)0.183	(±)0.094	(±)0.012
	$\|\mathcal{B}\|$	45.97	45.97	34	83
		(±)1.26	(±)1.26	(±)0	(±)2.89
	K	3.05	7.99	7.94	8
		(±)0.06	(±)0.02	(±)0.03	(±)0
	J_{intra}	6.213	2.881	3.855	2.974
		(±)0.07	(±)0.072	(±)0.075	(±)0.044
4	J_{inter}	21.548	19.806	18.53	19.939
		(±)0.11	(±)0.109	(±)0.248	(±)0.174
	$\hat{Q}$	0.332	0.635	1.269	0.426
		(±)0.02	(±)0.198	(±)0.073	(±)0.02
	$\|\mathcal{B}\|$	46.19	46.19	44	37.09
		(±)1.78	(±)1.78	(±)0	(±)2.45
	K	3.1	7.98	7.95	8
		(±)0.26	(±)0.03	(±)0.02	(±)0
	J_{intra}	6.033	2.85	3.908	2.845
		(±)0.24	(±)0.041	(±)0.062	(±)0.113
5	J_{inter}	21.636	19.877	18.558	20.187
		(±)0.185	(±)0.04	(±)0.171	(±)0.207

Continued on next page

$\tilde{f}$		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	$\hat{Q}$	0.323	0.601	1.274	0.437
		(±)0.028	(±)0.119	(±)0.048	(±)0.017
	$\|\mathcal{B}\|$	38.2	38.2	37	55.03
		(±)1.06	(±)1.06	(±)0	(±)2.9

Table 7.20: Descriptive Statistics: Pattern migration at different severities of change (N=8; $|C|$=25; $\tilde{f}$=3)

$\tilde{s}$		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	K	3.02	7.99	7.95	8
		(±)0.03	(±)0.02	(±)0.02	(±)0
	J_{intra}	6.067	2.829	3.945	2.762
		(±)0.052	(±)0.051	(±)0.149	(±)0.017
1	J_{inter}	21.822	19.963	18.71	20.045
		(±)0.065	(±)0.066	(±)0.271	(±)0.096
	$\hat{Q}$	0.314	0.594	1.263	0.435
		(±)0.012	(±)0.146	(±)0.096	(±)0.018
	$\|\mathcal{B}\|$	45.42	45.42	34	68.13
		(±)1.67	(±)1.67	(±)0	(±)2.96
	K	3.17	7.99	7.96	8
		(±)0.56	(±)0.02	(±)0.02	(±)0
	J_{intra}	5.931	2.862	3.88	2.776
		(±)0.371	(±)0.072	(±)0.063	(±)0.015
2	J_{inter}	21.765	19.898	18.605	20.072
		(±)0.438	(±)0.097	(±)0.173	(±)0.067
	$\hat{Q}$	0.346	0.6	1.26	0.453
		(±)0.125	(±)0.142	(±)0.041	(±)0.016
	$\|\mathcal{B}\|$	45.76	45.76	37	73.46
		(±)1.74	(±)1.74	(±)0	(±)2.73
	K	3.14	7.99	7.92	8
		(±)0.32	(±)0.01	(±)0.04	(±)0
	J_{intra}	5.997	2.863	4.004	2.793

Continued on next page

$\tilde{s}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
		(±)0.348	(±)0.07	(±)0.117	(±)0.014
3	J_{inter}	21.557	19.887	18.53	19.869
		(±)0.219	(±)0.068	(±)0.32	(±)0.084
	$\hat{Q}$	0.335	0.586	1.286	0.436
		(±)0.029	(±)0.183	(±)0.094	(±)0.012
	$\lvert\mathcal{B}\rvert$	45.97	45.97	34	83
		(±)1.26	(±)1.26	(±)0	(±)2.89
	K	3.06	7.99	7.9	8
		(±)0.36	(±)0.02	(±)0.05	(±)0
	J_{intra}	6.202	2.898	3.977	2.81
		(±)0.641	(±)0.057	(±)0.107	(±)0.015
4	J_{inter}	21.648	19.941	18.497	20.008
		(±)0.398	(±)0.096	(±)0.388	(±)0.093
	$\hat{Q}$	0.337	0.598	1.275	0.438
		(±)0.047	(±)0.122	(±)0.095	(±)0.014
	$\lvert\mathcal{B}\rvert$	45.7	45.7	41	78.81
		(±)1.38	(±)1.38	(±)0	(±)3.21
	K	3.23	7.99	7.95	8
		(±)0.87	(±)0.03	(±)0.02	(±)0
	J_{intra}	6.023	2.891	3.944	2.82
		(±)0.898	(±)0.045	(±)0.053	(±)0.015
5	J_{inter}	21.649	19.976	18.578	20.163
		(±)0.471	(±)0.076	(±)0.208	(±)0.068
	$\hat{Q}$	0.339	0.601	1.269	0.444
		(±)0.058	(±)0.119	(±)0.04	(±)0.011
	$\lvert\mathcal{B}\rvert$	45.88	45.88	37	78.66
		(±)1.67	(±)1.67	(±)0	(±)2.71

7.5.2 Cluster Migration

Figure 7.7 illustrates the quality of partitioning by the different models over time for cluster migration. Similar trends as for pattern migration are illustrated for cluster quality (see figure 7.7(a)) and ALC population sizes (see figure 7.7(d)). As for pattern migration there is an increase in the ALC population size for the SMAIN model with every change in the data (figure 7.7(d) at $t = 30$, $t = 60$ and $t = 90$). The drawback of SMAIN to potentially overfit the data is emphasised with cluster migration environments, since it is expected to utilise a smaller ALC population size with a decrease in the number of clusters in the data (as illustrated in figures 7.7(d) and 7.7(e), the ALC population size of SMAIN increases even with a decrease in the number of clusters). Figure 7.7(e) illustrates that LNNAIS and SMAIN detected the change in the number of clusters at $t = 30$. The expected number of clusters for $t \geq 30$ is $K = 6$ which is correctly obtained by LNNAIS and SMAIN. DWB tends to cluster the data into slightly more than six clusters ($6 < K < 7$), because of the hybrid approach followed by DWB (using K-means clustering). The DWB model partitions the ALC population into the initial eight clusters (eight sub-nets) at each step in time. This results into an average of 6.97 clusters at $t \geq 30$ (as illustrated in figure 7.7(e)), which explains the lower quality of clusters found by DWB when compared to SMAIN and LNNAIS (as illustrated in figure 7.7(a)). Again, LNN$_{SDOT}$ did not detect the correct number of clusters, but did however detect a change in the data at $t = 30$, $t = 60$ and $t = 90$ (see figure 7.7(f)). Overall, there is no significant change in Q_{ratio}, J_{intra}, J_{inter} and K for all the models ($t > 40$) in cluster migration environments except for LNN$_{SDOT}$ where K fluctuates between $K = 2$ and $K = 3$ at every change in the data (explained in section 7.5.3).

The Mann-Whitney U statistical hypothesis test rejects H_0 that the mean clustering quality, $\bar{Q}$, are the same at a 0.05 level of significance between LNNAIS and LNN$_{SDOT}$, SMAIN and DWB for different dimensions (as summarised in table 7.21), different clusters sizes (as summarised in table 7.22), for all frequencies of change (as summarised in table 7.23), and all severities of change (as summarised in table 7.24). There is thus a statistical significant difference in the clustering quality of all the cluster migration data sets between LNNAIS and all the other models.

Table 7.21: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Cluster migration at different dimensions ($|C|$=25; $\tilde{f}$=3; $\tilde{s}$=3)

N	LNN$_{SDOT}$	DWB	SMAIN
3	$z = 4.568$	$z = 6.646$	$z = 6.646$
3	$p < 0.001$	$p < 0.001$	$p < 0.001$
3	Reject H_0	Reject H_0	Reject H_0
8	$z = 6.646$	$z = 6.646$	$z = 6.646$

Continued on next page

N	LNN_{SDOT}	DWB	SMAIN
8	$p < 0.001$	$p < 0.001$	$p < 0.001$
8	Reject H_0	Reject H_0	Reject H_0
15	$z = 5.455$	$z = 5.862$	$z = 5.64$
15	$p < 0.001$	$p < 0.001$	$p < 0.001$
15	Reject H_0	Reject H_0	Reject H_0

Table 7.22: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Cluster migration at different cluster sizes (N=8; $\tilde{s}$=3; $\tilde{f}$=3)

$\|C\|$	LNN_{SDOT}	DWB	SMAIN
10	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.527$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
25	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
50	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 3.785$ $p < 0.001$ Reject H_0

Table 7.23: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Cluster migration at different frequencies of change (N=8; $|C|$=25; $\tilde{s}$=3)

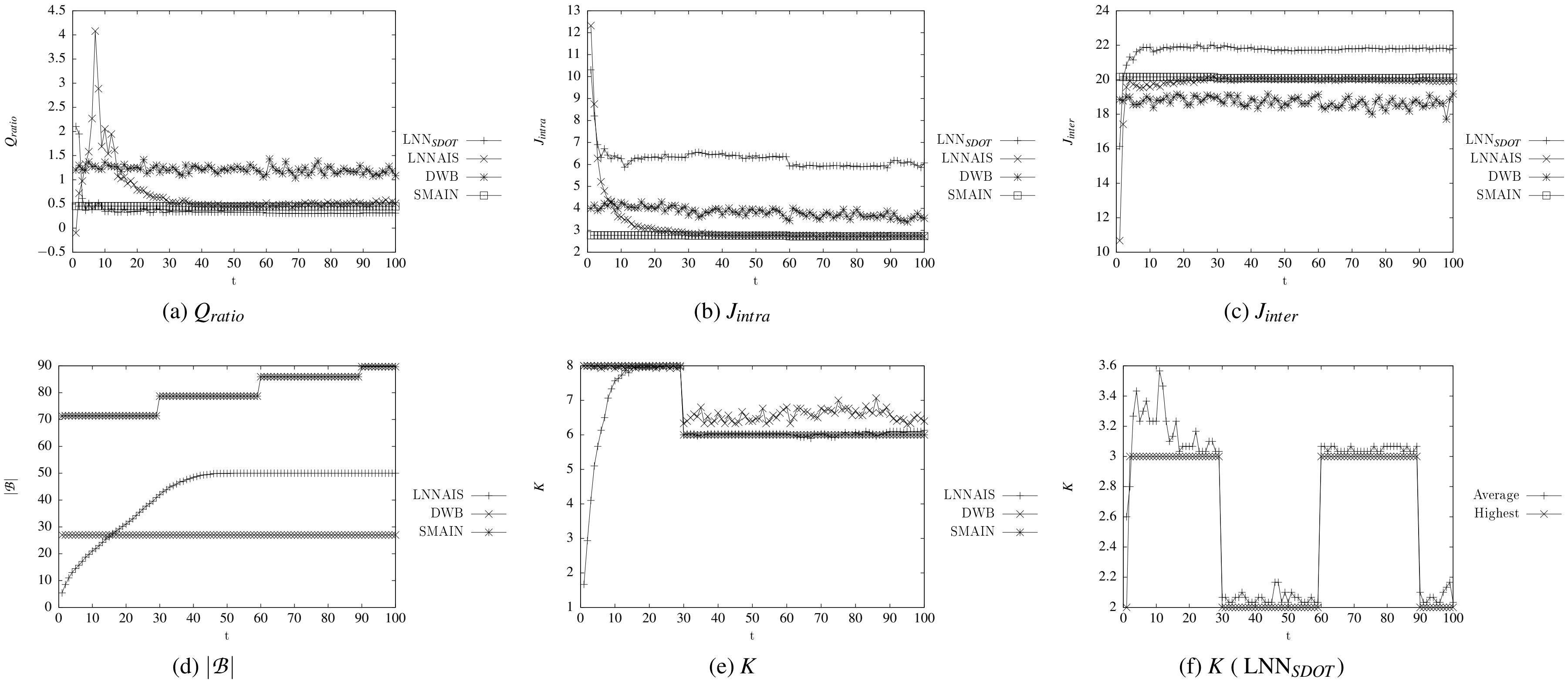
$\tilde{f}$	LNN_{SDOT}	DWB	SMAIN
1	$z = 6.631$ $p < 0.001$ Reject H_0	$z = 6.631$ $p < 0.001$ Reject H_0	$z = 6.631$ $p < 0.001$ Reject H_0
	$z = 5.877$	$z = 6.646$	$z = 6.616$

Continued on next page

$\tilde{f}$	LNN_{SDOT}	DWB	SMAIN
2	$p < 0.001$ Reject H_0	$p < 0.001$ Reject H_0	$p < 0.001$ Reject H_0
3	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
4	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.631$ $p < 0.001$ Reject H_0
5	$z = 6.35$ $p < 0.001$ Reject H_0	$z = 6.276$ $p < 0.001$ Reject H_0	$z = 6.579$ $p < 0.001$ Reject H_0

Table 7.24: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Cluster migration at different severities of change (N=8; $|C|$=25; $\tilde{f}$=3)

$\tilde{s}$	LNN_{SDOT}	DWB	SMAIN
1	$z = 6.291$ $p < 0.001$ Reject H_0	$z = 6.202$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
2	$z = 6.601$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
3	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
4	$z = 6.431$ $p < 0.001$ Reject H_0	$z = 6.527$ $p < 0.001$ Reject H_0	$z = 6.631$ $p < 0.001$ Reject H_0
5	$z = 6.328$ $p < 0.001$ Reject H_0	$z = 6.601$ $p < 0.001$ Reject H_0	$z = 6.32$ $p < 0.001$ Reject H_0



(a) Q_{ratio} (b) J_{intra} (c) J_{inter}

(d) $|\mathcal{B}|$ (e) K (f) K (LNN$_{SDOT}$)

Figure 7.7 Cluster Migration ($N = 8, |C| = 25, \tilde{f} = 3, \tilde{s} = 3$): Quantifying each model's partitioning quality with regards to Q_{ratio}, $|\mathcal{B}|$, J_{intra}, J_{inter} and K

Similar to the pattern migration environments, SMAIN utilises a larger ALC population size than LNNAIS and tends to overfit the data. This results in clusters of higher quality for cluster migration environments when compared to LNNAIS at different dimensions (see table 7.25), cluster sizes (see table 7.26), frequencies of change (see table 7.27) and severities of change (see table 7.28). The drawback of overfit is even more emphasised in cluster migration environments where the ALC population size of the SMAIN model does not scale with the number of clusters. LNNAIS delivers clusters of a higher quality than DWB and LNN_{SDOT} for all cluster migration environments. LNNAIS also obtains the correct number of clusters at different severities of change with no significant change in the ALC population size (see table 7.28 where an increase in $\tilde{s}$ increases the number of clusters migrating and disappearing in the data, i.e. decreasing the number of clusters in the data).

Table 7.25: Descriptive Statistics: Cluster migration at different dimensions ($|C|$=25; $\tilde{f}$=3; $\tilde{s}$=3)

N		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	K	4.38	6.46	7.25	6.58
		(±)1.31	(±)0.08	(±)0.08	(±)0
	J_{intra}	5.565	1.66	2.711	1.555
		(±)2.487	(±)0.132	(±)0.106	(±)0.008
3	J_{inter}	21.426	21.609	20.031	21.707
		(±)0.585	(±)0.147	(±)0.163	(±)0.053
	$\hat{Q}$	0.367	0.224	1.468	0.14
		(±)0.124	(±)0.102	(±)0.093	(±)0.002
	$\|\mathcal{B}\|$	49.15	49.15	95	83.15
		(±)0.47	(±)0.47	(±)0	(±)2.51
	K	2.62	6.43	6.97	6.58
		(±)0.18	(±)0.09	(±)0.2	(±)0
	J_{intra}	6.174	2.835	3.835	2.75
		(±)0.282	(±)0.046	(±)0.164	(±)0.015
8	J_{inter}	21.799	19.979	18.717	20.118
		(±)0.21	(±)0.085	(±)0.315	(±)0.074
	$\hat{Q}$	0.326	0.59	1.214	0.45
		(±)0.023	(±)0.105	(±)0.088	(±)0.011

Continued on next page

N		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	$\|\mathcal{B}\|$	45.49	45.49	27	79.97
		(±)1.63	(±)1.63	(±)0	(±)3.83
	K	3.69	3.8	7.45	6.64
		(±)0.11	(±)0.06	(±)0.14	(±)0.11
	J_{intra}	6.705	6.845	4.93	4.034
		(±)0.225	(±)0.163	(±)0.183	(±)0.083
15	J_{inter}	21.52	20.84	19.678	20.739
		(±)0.155	(±)0.165	(±)0.344	(±)0.169
	$\hat{Q}$	0.553	0.856	1.219	0.633
		(±)0.13	(±)0.188	(±)0.046	(±)0.019
	$\|\mathcal{B}\|$	10	10	31	79.7
		(±)0	(±)0	(±)0	(±)3.19

Table 7.26: Descriptive Statistics: Cluster migration at different cluster sizes (N=8; $\tilde{s}$=3; $\tilde{f}$=3)

$\|C\|$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	3.11	3.7	7.2	6.58
		(±)0.09	(±)0.05	(±)0.13	(±)0
	J_{intra}	6.046	5.389	3.671	2.752
		(±)0.185	(±)0.109	(±)0.117	(±)0.008
10	J_{inter}	21.989	20.648	18.821	20.097
		(±)0.252	(±)0.149	(±)0.343	(±)0.054
	$\hat{Q}$	0.386	0.816	1.132	0.421
		(±)0.045	(±)0.105	(±)0.067	(±)0.009
	$\|\mathcal{B}\|$	10	10	35	71.79
		(±)0.01	(±)0.01	(±)0	(±)1.11
	K	2.62	6.43	6.97	6.58
		(±)0.18	(±)0.09	(±)0.2	(±)0
	J_{intra}	6.174	2.835	3.835	2.75
		(±)0.282	(±)0.046	(±)0.164	(±)0.015
25	J_{inter}	21.799	19.979	18.717	20.118

Continued on next page

$\|C\|$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
		(±)0.21	(±)0.085	(±)0.315	(±)0.074
	$\hat{Q}$	0.326	0.59	1.214	0.45
		(±)0.023	(±)0.105	(±)0.088	(±)0.011
	$\|\mathcal{B}\|$	45.49	45.49	27	79.97
		(±)1.63	(±)1.63	(±)0	(±)3.83
	K	3.04	6.56	7.24	6.58
		(±)0.14	(±)0.13	(±)0.14	(±)0
	J_{intra}	5.565	2.851	3.808	2.864
		(±)0.231	(±)0.018	(±)0.108	(±)0.017
50	J_{inter}	21.576	19.938	18.27	20.021
		(±)0.153	(±)0.083	(±)0.44	(±)0.1
	$\hat{Q}$	0.295	0.522	1.236	0.452
		(±)0.023	(±)0.104	(±)0.068	(±)0.017
	$\|\mathcal{B}\|$	48.89	48.89	31	72
		(±)0.73	(±)0.73	(±)0	(±)3.05

Table 7.27: Descriptive Statistics: Cluster migration at different frequencies of change (N=8; $|C|$=25; $\tilde{s}$=3)

$\tilde{f}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	2.95	6.26	7.03	6.25
		(±)0.2	(±)0.12	(±)0.12	(±)0.05
	J_{intra}	5.617	2.833	3.617	2.822
		(±)0.472	(±)0.041	(±)0.067	(±)0.017
1	J_{inter}	21.667	19.686	18.281	20.307
		(±)0.173	(±)0.156	(±)0.357	(±)0.075
	$\hat{Q}$	0.303	0.77	1.203	0.445
		(±)0.044	(±)0.167	(±)0.062	(±)0.016
	$\|\mathcal{B}\|$	37.97	37.97	29	69.26
		(±)0.73	(±)0.73	(±)0	(±)2.75
	K	3.19	6.38	7.2	6.38
		(±)0.37	(±)0.15	(±)0.11	(±)0

Continued on next page

$\tilde{f}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	J_{intra}	5.764	2.832	3.681	2.793
		(±)0.613	(±)0.046	(±)0.078	(±)0.015
2	J_{inter}	21.561	19.94	18.429	20.145
		(±)0.679	(±)0.112	(±)0.323	(±)0.087
	$\hat{Q}$	0.366	0.671	1.217	0.439
		(±)0.11	(±)0.141	(±)0.062	(±)0.016
	$\|\mathcal{B}\|$	45.68	45.68	31	74.82
		(±)1.64	(±)1.64	(±)0	(±)2.84
	K	2.62	6.43	6.97	6.58
		(±)0.18	(±)0.09	(±)0.2	(±)0
	J_{intra}	6.174	2.835	3.835	2.75
		(±)0.282	(±)0.046	(±)0.164	(±)0.015
3	J_{inter}	21.799	19.979	18.717	20.118
		(±)0.21	(±)0.085	(±)0.315	(±)0.074
	$\hat{Q}$	0.326	0.59	1.214	0.45
		(±)0.023	(±)0.105	(±)0.088	(±)0.011
	$\|\mathcal{B}\|$	45.49	45.49	27	79.97
		(±)1.63	(±)1.63	(±)0	(±)3.83
	K	2.95	6.68	7.61	6.78
		(±)0.45	(±)0.12	(±)0.13	(±)0
	J_{intra}	5.691	2.895	3.702	2.783
		(±)0.477	(±)0.046	(±)0.063	(±)0.022
4	J_{inter}	21.621	19.878	18.223	19.944
		(±)0.372	(±)0.115	(±)0.205	(±)0.064
	$\hat{Q}$	0.324	0.658	1.221	0.455
		(±)0.054	(±)0.136	(±)0.038	(±)0.015
	$\|\mathcal{B}\|$	45.96	45.96	32	83.78
		(±)1.6	(±)1.6	(±)0	(±)2.24
	K	3.03	6.96	7.45	6.98
		(±)0.28	(±)0.16	(±)0.24	(±)0
	J_{intra}	6.086	2.932	3.965	2.828
		(±)0.728	(±)0.079	(±)0.129	(±)0.012

Continued on next page

$\tilde{f}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
5	J_{inter}	21.726	19.77	18.519	20.023
		(±)0.295	(±)0.181	(±)0.405	(±)0.095
	$\hat{Q}$	0.337	0.695	1.215	0.439
		(±)0.076	(±)0.208	(±)0.079	(±)0.012
	$\|\mathcal{B}\|$	46.05	46.05	27	80.23
		(±)1.77	(±)1.77	(±)0	(±)3.46

Table 7.28: Descriptive Statistics: Cluster migration at different severities of change (N=8; $|C|$=25; $\tilde{f}$=3)

$\tilde{s}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	3.02	7.99	7.95	8
		(±)0.26	(±)0.02	(±)0.02	(±)0
	J_{intra}	6.336	2.892	3.908	2.807
		(±)0.69	(±)0.067	(±)0.103	(±)0.013
1	J_{inter}	21.911	20.105	18.692	20.178
		(±)0.303	(±)0.081	(±)0.251	(±)0.062
	$\hat{Q}$	0.342	0.609	1.262	0.422
		(±)0.065	(±)0.168	(±)0.063	(±)0.017
	$\|\mathcal{B}\|$	46.07	46.07	44	73.39
		(±)1.53	(±)1.53	(±)0	(±)2.82
	K	3.04	7.26	7.75	7.29
		(±)0.1	(±)0.18	(±)0.09	(±)0
	J_{intra}	6.118	2.747	3.722	2.705
		(±)0.251	(±)0.06	(±)0.072	(±)0.02
2	J_{inter}	21.617	19.827	18.552	20.051
		(±)0.228	(±)0.193	(±)0.201	(±)0.078
	$\hat{Q}$	0.328	0.616	1.27	0.384
		(±)0.027	(±)0.157	(±)0.047	(±)0.01
	$\|\mathcal{B}\|$	45.84	45.84	36	69.04
		(±)1.32	(±)1.32	(±)0	(±)3.22

Continued on next page

$\tilde{s}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	2.62	6.43	6.97	6.58
		$(\pm)0.18$	$(\pm)0.09$	$(\pm)0.2$	$(\pm)0$
	J_{intra}	6.174	2.835	3.835	2.75
		$(\pm)0.282$	$(\pm)0.046$	$(\pm)0.164$	$(\pm)0.015$
3	J_{inter}	21.799	19.979	18.717	20.118
		$(\pm)0.21$	$(\pm)0.085$	$(\pm)0.315$	$(\pm)0.074$
	$\hat{Q}$	0.326	0.59	1.214	0.45
		$(\pm)0.023$	$(\pm)0.105$	$(\pm)0.088$	$(\pm)0.011$
	$\lvert\mathcal{B}\rvert$	45.49	45.49	27	79.97
		$(\pm)1.63$	$(\pm)1.63$	$(\pm)0$	$(\pm)3.83$
	K	3.06	5.75	7.29	5.87
		$(\pm)0.11$	$(\pm)0.15$	$(\pm)0.07$	$(\pm)0$
	J_{intra}	5.787	2.81	3.448	2.79
		$(\pm)0.398$	$(\pm)0.036$	$(\pm)0.068$	$(\pm)0.017$
4	J_{inter}	21.516	19.729	17.926	19.754
		$(\pm)0.356$	$(\pm)0.171$	$(\pm)0.231$	$(\pm)0.11$
	$\hat{Q}$	0.322	0.689	1.207	0.425
		$(\pm)0.058$	$(\pm)0.211$	$(\pm)0.044$	$(\pm)0.014$
	$\lvert\mathcal{B}\rvert$	45.6	45.6	36	63.18
		$(\pm)1.64$	$(\pm)1.64$	$(\pm)0$	$(\pm)2.33$
	K	3.11	4.86	6.4	5.13
		$(\pm)0.29$	$(\pm)0.08$	$(\pm)0.11$	$(\pm)0.09$
	J_{intra}	5.634	2.863	3.473	2.879
		$(\pm)0.575$	$(\pm)0.042$	$(\pm)0.074$	$(\pm)0.088$
5	J_{inter}	21.489	19.832	18.296	19.861
		$(\pm)0.637$	$(\pm)0.109$	$(\pm)0.272$	$(\pm)0.148$
	$\hat{Q}$	0.333	0.615	1.176	0.437
		$(\pm)0.07$	$(\pm)0.142$	$(\pm)0.059$	$(\pm)0.021$
	$\lvert\mathcal{B}\rvert$	45.49	45.49	39	65.09
		$(\pm)1.45$	$(\pm)1.45$	$(\pm)0$	$(\pm)2.89$

7.5.3 Centroid Migration

Figure 7.8 illustrates the quality of partitioning by the different models over time for centroid migration. Different to the pattern and cluster migration data sets, the centroids in the centroid migration data sets are non-stationary. Non-stationary centroids result in merging of clusters and division of clusters. Figure 7.8(e) illustrates that LNNAIS and SMAIN detected the change in the number of clusters at $t = 30$, $t = 60$ and $t = 90$. The expected number of clusters for $30 \leq t < 60$ and $60 \leq t < 90$ is $K = 6$ and $K = 4$ for $t \geq 90$ which is correctly obtained by LNNAIS and SMAIN. A similar drawback of DWB for cluster migration as for pattern and cluster migration is that DWB tends to cluster the data into slightly more clusters, because of the hybrid approach followed by DWB (using K-means clustering). The DWB model partitions the ALC population into the initial eight clusters (eight sub-nets) at each step in time. Different to pattern and cluster migration, LNN_{SDOT} detected the change in the data at $t = 30$, $t = 60$ and $t = 90$ and determined the correct number of clusters. LNN_{SDOT} did not determine the correct number of clusters for $t < 30$. Since the centroids in the centroid migration is non-stationary, the distances between these centroids change over time. This has a direct influence on the network affinity between the ALCs in LNNAIS, since the ALCs adapt to the clusters. As a result of the changes in the distance between the centroids, the sequential outlier technique detects more network affinities as outliers (utilised by LNN_{SDOT} to dynamically determine the ALC network boundaries, as explained in section 6.2). Detecting more ALC network boundaries correctly determines the number of clusters in the data set. This highlights a potential drawback of the sequential deviation outlier detection technique used by LNN_{SDOT} which is that if the centroids of clusters in a data set are uniformly distributed with equal distances, the ALC networks formed in LNNAIS will have equal network affinities between each other, resulting in no outlier network affinities. This is expected since the sequential deviation outlier detection technique will detect no outliers and therefore no boundaries between the ALC networks. This was the case for the pattern and cluster migration environments where the spatial positions of the centroids remain stationary (refer to figures 7.6(f) and figures 7.7(f) where $K \leq 3$ at all time steps). The same argument applies to centroid migration where $t < 30$ as illustrated in figure 7.8(f), since prior to this point in time none of the centroids have changed their spatial positions and only three boundaries between the ALC networks were detected by LNN_{SDOT}.

The drawback of SMAIN to potentially overfit the data is also highlighted with centroid migration environments, since it is expected to utilise a smaller ALC population size with a decrease in the number of clusters in the data (as illustrated in figures 7.8(d), 7.8(e) and 7.8(a); the

ALC population size of SMAIN increases even at a decrease in the number of clusters with no significant gain in the cluster quality). Furthermore the clusters found by SMAIN become less compact at each change (as illustrated in figure 7.8(b)). Figures 7.8(a) shows that the quality of clusters found by LNNAIS lowers at each change, but that LNNAIS succeeds to recover from the change and improve on the cluster quality as time progresses, even though the number of clusters changes and clusters become less compact (as illustrated in figures 7.8(b) and 7.8(e)). LNNAIS delivers clusters of a higher quality than DWB (for all t) and similar quality as LNN_{SDOT} (for $t > 30$) for all centroid migration environments.

Table 7.29: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Centroid migration at different dimensions ($|C|$=25; $\tilde{f}$=3; $\tilde{s}$=3)

N	LNN_{SDOT}	DWB	SMAIN
3	$z = 3.326$	$z = 6.646$	$z = 5.914$
3	$p < 0.001$	$p < 0.001$	$p < 0.001$
3	Reject H_0	Reject H_0	Reject H_0
8	$z = 4.805$	$z = 6.646$	$z = 6.646$
8	$p < 0.001$	$p < 0.001$	$p < 0.001$
8	Reject H_0	Reject H_0	Reject H_0
15	$z = 6.498$	$z = 6.646$	$z = 6.276$
15	$p < 0.001$	$p < 0.001$	$p < 0.001$
15	Reject H_0	Reject H_0	Reject H_0

Table 7.30: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Centroid migration at different cluster sizes (N=8; $\tilde{s}$=3; $\tilde{f}$=3)

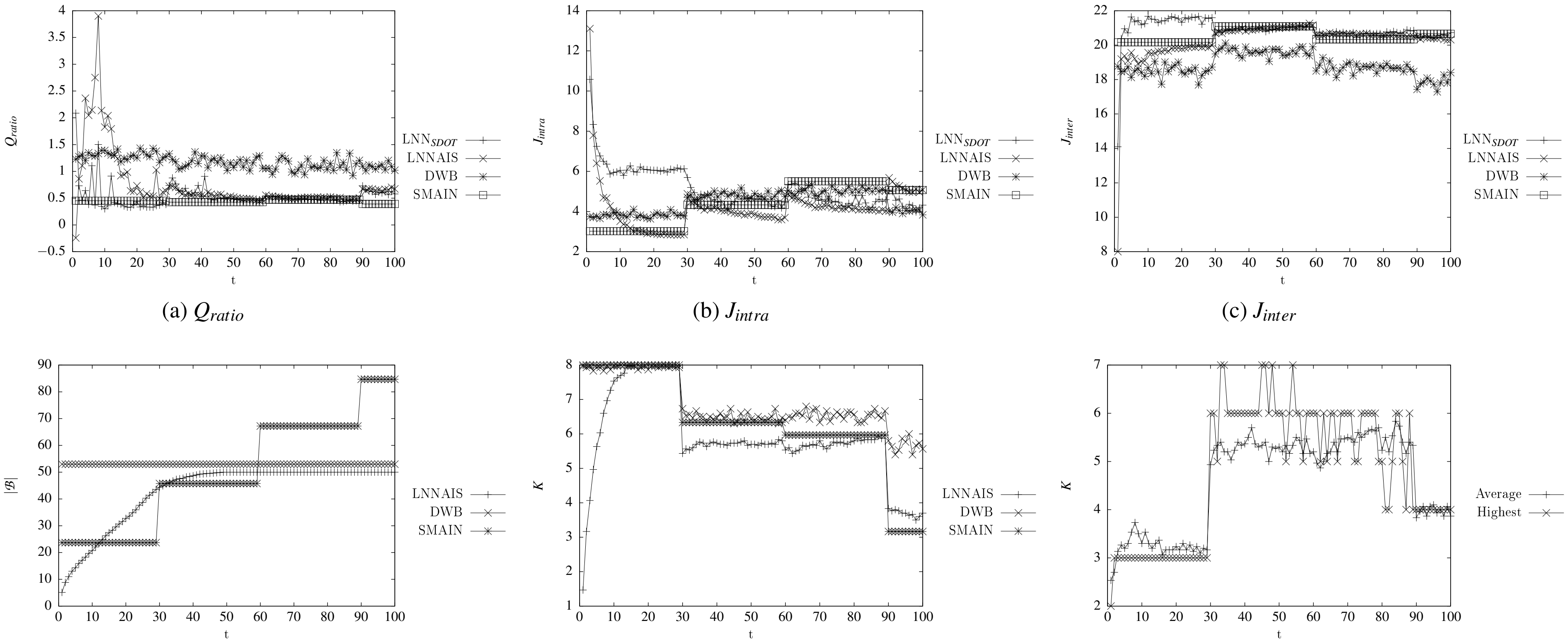
$\|C\|$	LNN_{SDOT}	DWB	SMAIN
	$z = 6.232$	$z = 6.209$	$z = 6.646$
10	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Reject H_0	Reject H_0	Reject H_0
	$z = 4.805$	$z = 6.646$	$z = 6.646$

Continued on next page

$\|C\|$	LNN_{SDOT}	DWB	SMAIN
25	$p < 0.001$ Reject H_0	$p < 0.001$ Reject H_0	$p < 0.001$ Reject H_0
50	$z = 5.67$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 4.709$ $p < 0.001$ Reject H_0

Table 7.31: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Centroid migration at different frequencies of change (N=8; $|C|$=25; $\tilde{s}$=3)

$\tilde{f}$	LNN_{SDOT}	DWB	SMAIN
1	$z = 3.674$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
2	$z = 6.158$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.631$ $p < 0.001$ Reject H_0
3	$z = 4.805$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
4	$z = 4.687$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 5.722$ $p < 0.001$ Reject H_0
5	$z = 5.241$ $p < 0.001$ Reject H_0	$z = 6.453$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0



(a) Q_{ratio} (b) J_{intra} (c) J_{inter}

(d) $|\mathcal{B}|$ (e) K (f) K (LNN$_{SDOT}$)

Figure 7.8 Centroid Migration ($N = 8, |C| = 25, \tilde{f} = 3, \tilde{s} = 3$): Quantifying each model's partitioning quality with regards to Q_{ratio}, $|\mathcal{B}|$, J_{intra}, J_{inter} and K

Table 7.32: Statistical Hypothesis Testing between LNNAIS and Other Models ($\alpha = 0.05$; with continuity correction; unpaired; non-directional) for Centroid migration at different severities of change (N=8; $|C|$=25; $\tilde{f}$=3)

$\tilde{s}$	LNN_{SDOT}	DWB	SMAIN
1	$z = 5.345$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 0.074$ $p = 0.941$ Accept H_0
2	$z = 5.98$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.527$ $p < 0.001$ Reject H_0
3	$z = 4.805$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0
4	$z = 4.273$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 5.781$ $p < 0.001$ Reject H_0
5	$z = 4.938$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0	$z = 6.646$ $p < 0.001$ Reject H_0

The Mann-Whitney U statistical hypothesis test rejects H_0 that the mean clustering quality, $\bar{Q}$, are the same at a 0.05 level of significance between LNNAIS and LNN_{SDOT}, SMAIN and DWB for different dimensions (as summarised in table 7.29), different clusters sizes (as summarised in table 7.30), for all frequencies of change (as summarised in table 7.31) and severities of change greater than one (as summarised in table 7.32). There is thus a statistical significant difference in the clustering quality of all the centroid migration data sets between LNNAIS and all the other models except for $\tilde{s} = 1$ between LNNAIS and SMAIN where the Mann-Whitney U statistical hypothesis test accepted H_0.

Table 7.33: Descriptive Statistics: Centroid migration at different dimensions ($|C|$=25; $\tilde{f}$=3; $\tilde{s}$=3)

N		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	3.4	5.69	6.73	5.63
		(±)0.84	(±)0.25	(±)0.12	(±)0.18
	J_{intra}	7.111	2.49	2.949	2.694
		(±)2.193	(±)0.246	(±)0.078	(±)0.163
3	J_{inter}	20.43	21.022	19.256	21.102
		(±)1.01	(±)0.3	(±)0.22	(±)0.127
	$\hat{Q}$	0.493	0.392	1.238	0.253
		(±)0.1	(±)0.121	(±)0.067	(±)0.015
	$\lvert\mathcal{B}\rvert$	49.18	49.18	95	60.01
		(±)0.62	(±)0.62	(±)0	(±)2.13
	K	4.73	5.94	6.84	6.36
		(±)0.55	(±)0.18	(±)0.08	(±)0.2
	J_{intra}	4.919	4.01	4.5	4.386
		(±)0.621	(±)0.198	(±)0.093	(±)0.11
8	J_{inter}	20.933	20.522	18.828	20.548
		(±)0.316	(±)0.196	(±)0.245	(±)0.136
	$\hat{Q}$	0.507	0.62	1.18	0.443
		(±)0.101	(±)0.092	(±)0.065	(±)0.011
	$\lvert\mathcal{B}\rvert$	45.97	45.97	53	50.09
		(±)1.6	(±)1.6	(±)0	(±)2.82
	K	5.56	5.98	6.66	6.06
		(±)0.31	(±)0.41	(±)0.18	(±)0.38
	J_{intra}	4.973	4.885	5.639	7.728
		(±)0.258	(±)0.315	(±)0.118	(±)1.248
15	J_{inter}	20.601	20.069	18.641	20.45
		(±)0.29	(±)0.51	(±)0.281	(±)0.775
	$\hat{Q}$	0.529	0.772	1.184	1.28
		(±)0.056	(±)0.114	(±)0.064	(±)0.43
	$\lvert\mathcal{B}\rvert$	28.94	28.94	68	28.22
		(±)0.64	(±)0.64	(±)0	(±)3.32

Table 7.34: Descriptive Statistics: Centroid migration at different cluster sizes (N=8; $\tilde{s}$=3; $\tilde{f}$=3)

$\|C\|$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	3.05	3.62	6.81	5.73
		(±)0.1	(±)0.09	(±)0.13	(±)0.14
	J_{intra}	7.143	6.531	4.562	4.61
		(±)0.178	(±)0.175	(±)0.084	(±)0.091
10	J_{inter}	21.198	19.536	18.374	19.971
		(±)0.252	(±)0.257	(±)0.222	(±)0.137
	$\hat{Q}$	0.511	0.894	1.131	0.479
		(±)0.113	(±)0.107	(±)0.048	(±)0.014
	$\|\mathcal{B}\|$	10	10	53	45.68
		(±)0.01	(±)0.01	(±)0	(±)2.22
	K	4.73	5.94	6.84	6.36
		(±)0.55	(±)0.18	(±)0.08	(±)0.2
	J_{intra}	4.919	4.01	4.5	4.386
		(±)0.621	(±)0.198	(±)0.093	(±)0.11
25	J_{inter}	20.933	20.522	18.828	20.548
		(±)0.316	(±)0.196	(±)0.245	(±)0.136
	$\hat{Q}$	0.507	0.62	1.18	0.443
		(±)0.101	(±)0.092	(±)0.065	(±)0.011
	$\|\mathcal{B}\|$	45.97	45.97	53	50.09
		(±)1.6	(±)1.6	(±)0	(±)2.82
	K	4.62	5.79	6.75	5.37
		(±)0.46	(±)0.26	(±)0.13	(±)0.19
	J_{intra}	4.774	3.958	4.492	4.532
		(±)0.593	(±)0.218	(±)0.087	(±)0.119
50	J_{inter}	21.195	20.542	18.53	20.706
		(±)0.283	(±)0.212	(±)0.321	(±)0.144
	$\hat{Q}$	0.441	0.542	1.241	0.489
		(±)0.05	(±)0.062	(±)0.049	(±)0.021
	$\|\mathcal{B}\|$	39.72	39.72	53	59.29
		(±)0.18	(±)0.18	(±)0	(±)2.83

Table 7.35: Descriptive Statistics: Centroid migration at different frequencies of change (N=8; $|C|$=25; $\tilde{s}$=3)

$\tilde{f}$		LNN_{SDOT}	LNNAIS	DWB	SMAIN
	K	5.28	5.66	5.88	5.24
		(±)0.46	(±)0.33	(±)0.17	(±)0.32
	J_{intra}	4.588	4.154	5.012	7.429
		(±)0.618	(±)0.254	(±)0.099	(±)0.437
1	J_{inter}	19.49	19.046	18.411	20.122
		(±)0.741	(±)0.54	(±)0.312	(±)0.615
	$\hat{Q}$	0.55	0.636	1.131	1.071
		(±)0.084	(±)0.096	(±)0.056	(±)0.172
	$\|\mathcal{B}\|$	29.5	29.51	95	29.31
		(±)0.4	(±)0.39	(±)0	(±)2.43
	K	4.51	5.36	6.3	4.81
		(±)0.35	(±)0.34	(±)0.14	(±)0.36
	J_{intra}	4.815	4.128	4.591	5
		(±)0.381	(±)0.301	(±)0.098	(±)0.289
2	J_{inter}	20.159	19.389	17.782	20.035
		(±)0.45	(±)0.691	(±)0.355	(±)0.452
	$\hat{Q}$	0.483	0.65	1.178	0.455
		(±)0.073	(±)0.105	(±)0.046	(±)0.019
	$\|\mathcal{B}\|$	29.51	29.52	53	56.85
		(±)0.36	(±)0.35	(±)0	(±)5.91
	K	4.73	5.94	6.84	6.36
		(±)0.55	(±)0.18	(±)0.08	(±)0.2
	J_{intra}	4.919	4.01	4.5	4.386
		(±)0.621	(±)0.198	(±)0.093	(±)0.11
3	J_{inter}	20.933	20.522	18.828	20.548
		(±)0.316	(±)0.196	(±)0.245	(±)0.136
	$\hat{Q}$	0.507	0.62	1.18	0.443
		(±)0.101	(±)0.092	(±)0.065	(±)0.011
	$\|\mathcal{B}\|$	45.97	45.97	53	50.09
		(±)1.6	(±)1.6	(±)0	(±)2.82

Continued on next page

$\tilde{f}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	5.56	6.69	7.09	6.82
		(±)0.35	(±)0.26	(±)0.13	(±)0.3
	J_{intra}	4.535	3.807	5.03	3.949
		(±)0.434	(±)0.243	(±)0.103	(±)0.389
4	J_{inter}	21.091	20.623	19.153	20.657
		(±)0.201	(±)0.129	(±)0.215	(±)0.208
	$\hat{Q}$	0.448	0.544	1.251	0.415
		(±)0.134	(±)0.106	(±)0.069	(±)0.028
	$\lvert\mathcal{B}\rvert$	38.18	38.18	40	32.22
		(±)0.72	(±)0.72	(±)0	(±)4.88
	K	4.71	7.38	7.34	7.47
		(±)0.75	(±)0.24	(±)0.08	(±)0.09
	J_{intra}	5.581	3.49	4.634	3.521
		(±)0.919	(±)0.185	(±)0.074	(±)0.114
5	J_{inter}	21.024	20.229	18.954	20.36
		(±)0.209	(±)0.253	(±)0.159	(±)0.12
	$\hat{Q}$	0.469	0.652	1.203	0.42
		(±)0.096	(±)0.144	(±)0.044	(±)0.018
	$\lvert\mathcal{B}\rvert$	38.12	38.12	68	51.92
		(±)0.58	(±)0.58	(±)0	(±)3.76

Table 7.36: Descriptive Statistics: Centroid migration at different severities of change (N=8; $|C|$=25; $\tilde{f}$=3)

$\tilde{s}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
	K	4.59	7.81	7.47	7.35
		(±)0.86	(±)0.21	(±)0.11	(±)0.12
	J_{intra}	6.045	3.162	4.934	4.121
		(±)1.106	(±)0.143	(±)0.097	(±)0.083
1	J_{inter}	21.012	20.319	18.815	20.031
		(±)0.345	(±)0.139	(±)0.24	(±)0.182
	$\hat{Q}$	0.409	0.523	1.288	0.496

Continued on next page

$\tilde{s}$		LNN_{SDOT}	LNNAIS	DWB	SMAIN
		(±)0.031	(±)0.096	(±)0.066	(±)0.02
	$\|\mathcal{B}\|$	37.89	37.89	46	46.68
		(±)0.95	(±)0.95	(±)0	(±)2.43
	K	5.68	6.53	6.91	6.02
		(±)0.59	(±)0.31	(±)0.07	(±)0.35
	J_{intra}	4.576	3.806	4.842	4.441
		(±)0.677	(±)0.232	(±)0.097	(±)0.288
2	J_{inter}	20.175	19.618	18.485	19.986
		(±)0.38	(±)0.307	(±)0.177	(±)0.264
	$\hat{Q}$	0.461	0.623	1.21	0.456
		(±)0.054	(±)0.086	(±)0.047	(±)0.024
	$\|\mathcal{B}\|$	37.81	37.81	69	39.93
		(±)0.91	(±)0.91	(±)0	(±)3.75
	K	4.73	5.94	6.84	6.36
		(±)0.55	(±)0.18	(±)0.08	(±)0.2
	J_{intra}	4.919	4.01	4.5	4.386
		(±)0.621	(±)0.198	(±)0.093	(±)0.11
3	J_{inter}	20.933	20.522	18.828	20.548
		(±)0.316	(±)0.196	(±)0.245	(±)0.136
	$\hat{Q}$	0.507	0.62	1.18	0.443
		(±)0.101	(±)0.092	(±)0.065	(±)0.011
	$\|\mathcal{B}\|$	45.97	45.97	53	50.09
		(±)1.6	(±)1.6	(±)0	(±)2.82
	K	5.38	5.92	6.66	5.88
		(±)0.51	(±)0.21	(±)0.1	(±)0.24
	J_{intra}	4.778	4.422	5.043	5.282
		(±)0.61	(±)0.219	(±)0.099	(±)0.27
4	J_{inter}	20.589	20.286	18.759	20.661
		(±)0.337	(±)0.271	(±)0.24	(±)0.18
	$\hat{Q}$	0.491	0.585	1.204	0.466
		(±)0.097	(±)0.091	(±)0.06	(±)0.025
	$\|\mathcal{B}\|$	38.42	38.42	69	47.99

Continued on next page

$\tilde{s}$		LNN$_{SDOT}$	LNNAIS	DWB	SMAIN
		(±)0.72	(±)0.72	(±)0	(±)2.84
	K	4.91	6.1	6.59	5.67
		(±)0.45	(±)0.25	(±)0.14	(±)0.15
	J_{intra}	4.586	3.728	4.743	4.305
		(±)0.426	(±)0.187	(±)0.075	(±)0.15
5	J_{inter}	20.852	20.309	18.845	20.672
		(±)0.348	(±)0.24	(±)0.252	(±)0.122
	$\hat{Q}$	0.503	0.635	1.222	0.441
		(±)0.076	(±)0.09	(±)0.046	(±)0.015
	$\lvert\mathcal{B}\rvert$	38.15	38.15	55	56.85
		(±)0.78	(±)0.78	(±)0	(±)3.54

Again, due to the tendency of SMAIN to overfit the data, the quality of clusters found by SMAIN for centroid migration environments generally tends to be higher than the quality of the clusters found by LNNAIS (see tables 7.33 - 7.36 for centroid migration environments with different dimensions, clusters sizes, frequencies of change and severities of change). Note that in cases where the ALC population of SMAIN has a similar size as the ALC population of LNNAIS, LNNAIS tends to deliver cluster of a higher quality than SMAIN. This is shown in table 7.33 for $N = 15$ where $|\mathcal{B}| \approx 28$ and $K \approx 6$ for LNN$_{SDOT}$, LNNAIS and SMAIN. Note that with these parameter values LNNAIS tends to deliver clusters with a higher quality than SMAIN and LNN$_{SDOT}$ tends to deliver clusters of a higher quality than both SMAIN and LNNAIS. The advantage of LNN$_{SDOT}$ is that the clusters were dynamically determined. This is also shown in table 7.35 for $\tilde{f} = 1$ where $|\mathcal{B}| \approx 29$ and $K \approx 5$ for LNN$_{SDOT}$, LNNAIS and SMAIN. LNNAIS tends to deliver clusters with a higher quality than SMAIN and LNN$_{SDOT}$ tends to deliver clusters of a higher quality than both SMAIN and LNNAIS. In general, LNNAIS delivers clusters of a higher quality than DWB for all centroid migration environments. LNNAIS also obtains the correct number of clusters at different severities of change with no significant change in the ALC population size (see table 7.36 where an increase in $\tilde{s}$ increases the ratio of centroid migration, i.e. decreasing the number of clusters in the data). In general, where LNN$_{SDOT}$ obtained the correct number of clusters, the quality of the clusters tends to be higher than those clusters delivered by LNNAIS at different dimensions, cluster sizes, frequencies of change and severities of change (see tables 7.33- 7.36). Furthermore, as discussed above, LNN$_{SDOT}$ also tends to deliver clusters of higher quality than SMAIN and DWB in cases where the data is not overfitted by SMAIN and a similar number of clusters are obtained.

7.6 Conclusion

The chapter discussed and investigated different data migration types in a non-stationary environment. These migration types were pattern migration, cluster migration and centroid migration. A procedure to generate artificial non-stationary data sets with different environment parameters and migration types was proposed. Also, clustering performance measures for a non-stationary environment were proposed. The proposed clustering performance measures were used for comparison between four network based artificial immune system models for clustering of the generated artificial non-stationary data sets. These models were LNNAIS, LNN_{SDOT}, DWB and SMAIN. A sensitivity analysis of the LNNAIS parameters was done on the different artificial non-stationary data sets for each of the defined data migration types.

A sensitivity analysis of the LNNAIS parameters shows that for all migration types, LNNAIS utilises small population sizes with small cluster sizes and larger population sizes for large cluster sizes. There is also no effect on $\mathcal{B}_{max}$ with different frequencies or severities of change. The clustering quality of LNNAIS is the lowest at high frequencies and high severities of change for all of the migration environments at different dimensions and cluster sizes. The clustering quality of LNNAIS improves with an increase in the cluster size at different dimensions. Increasing the number of dimensions lowers the clustering quality of LNNAIS at different cluster sizes. A difference between the migration types is that LNNAIS utilises small and large population sizes at different dimensions for centroid migration environments; but, for the other migration types LNNAIS utilises small population sizes for high dimensional environments with small cluster sizes. Also, the frequency and severity of change in high dimensional centroid migration environments have a smaller effect on the clustering performance of LNNAIS when compared to pattern and cluster migration environments.

Overall, the SMAIN model tends to find clusters of a higher quality for all types of data migration environments (at the cost of overfitting the data), followed by LNNAIS. The higher quality of clusters found by SMAIN is due to a larger ALC population size which is utilised by SMAIN. The drawback of overfit by SMAIN is even more emphasised in cluster and centroid migration environments where the ALC population size of the SMAIN model does not scale with the number of clusters, since it is expected to utilise a smaller ALC population size with a decrease in the number of clusters in the data. LNNAIS delivers clusters of a higher quality than DWB and LNN_{SDOT} for all pattern and cluster migration environments. In centroid migration environments, LNNAIS succeeds to recover from any changes and improve on the cluster quality as time progresses, even though the number of clusters changes and clusters become less compact. LNNAIS delivers clusters of a higher quality than DWB for centroid migration environments and in cases where the ALC population of SMAIN has a similar size as the ALC population of LNNAIS, LNNAIS also tends to deliver clusters of a higher quality than SMAIN. LNNAIS also obtains the correct number of clusters at different severities of change with no significant change in the ALC population size. Wherever LNN_{SDOT} obtained the correct number of clusters, the quality of the clusters tends to be higher than those clusters delivered by LNNAIS at different dimensions, clusters sizes, frequencies of change and severities of change. Furthermore, LNN_{SDOT} also tends to deliver clusters of higher quality than SMAIN and DWB in cases where the data is not overfitted by SMAIN and a similar number of clusters are obtained.

An advantage of LNNAIS and LNN_{SDOT}, compared to the other models, is that both models have less user specified parameters and are computationally less expensive since neither follows a hybrid approach like SMAIN and DWB to determine the number of ALC networks. A further advantage of LNN_{SDOT} is that the clusters are dynamically determined. A drawback of the SMAIN model in non-stationary environments is the increase in the ALC population size with each change in the data. This drawback has a major impact on the scalability of the SMAIN model. A drawback of the LNN_{SDOT} model is in cases where there are no outlier network affinities between ALC networks. The lack in outlier network affinities results in less network boundaries and therefore less ALC network (cluster) formations.

From the results presented, it can be concluded that LNNAIS, having a small set of control parameters, is an efficient clustering model for different non-stationary environments. LNN_{SDOT} is most suitable for centroid migration environments where the number of clusters in the data is not known and needs to be quantified over time.

Chapter 8

Conclusion

This chapter briefly highlights the findings and contributions of this thesis and discusses directions for future research.

8.1 Summary

This thesis investigated the application of a network theory inspired artificial immune model to data clustering problems in stationary and non-stationary environments.

Chapter 5 presented a new network based artificial immune model, namely the local network neighbourhood AIS (LNNAIS). The proposed model utilises an index based network topology to determine the network connectivity between the artificial lymphocytes (ALCs). The application of LNNAIS to data clustering problems in stationary environments was investigated. The clustering performance of the LNNAIS model was compared against classical clustering algorithms (K-means clustering and CPSO) and existing network based AIS models (SMAIN, DWB and Opt-aiNet). In most cases, LNNAIS produced better or similar results with reference to the clustering quality, compactness and separation of the clusters. Although SMAIN tends to deliver clusters of a higher quality than LNNAIS, further investigation into the size of the ALC populations showed that SMAIN utilised a larger ALC population to cluster the data. This explained the superior clustering quality of SMAIN but also highlighted a potential drawback of SMAIN that tend to overfit the data. Compared to SMAIN in view of these findings, the LNNAIS model delivers clusters of high quality without overfitting the data. A sensitivity analysis was done on the parameters of LNNAIS. The results suggest that an increase in the ALC population size increases diversity which obtains the required number of clusters and improves the clustering quality. Smaller neighbourhood sizes deliver more compact and more separated clusters when compared to larger neighbourhood sizes, and also tend to obtain the required number of clusters. Therefore small neighbourhood sizes deliver clusters of a higher quality. Furthermore, the clonal level threshold influences the compactness of the clusters and is problem specific.

Chapter 6 presented two different techniques which can be used by LNNAIS to dynamically determine the number clusters in a data set. These techniques are the iterative pruning technique (IPT) and the sequential deviation outlier technique (SDOT). Both of these techniques are computationally less expensive than the multiple execution approaches to dynamically determine the number of clusters in a data set. The IPT technique is computationally slightly more expensive than SDOT since IPT needs to iterate through all possible edges (to a maximum of $\mathcal{B}_{max}$). A range for K can be specified, but this makes IPT parameter dependant. An advantage of IPT is that the technique can use any cluster validity index to determine the number of clusters. The SDOT technique neither uses a cluster validity index nor does it require any boundary constraints on K. SDOT is a non-parametric technique. This is an advantage, since it is not always feasible to visually inspect the formed clusters and a specified range for K might not contain the optimum number of clusters. Both techniques were applied on different data sets to determine the optimal number of clusters. These results were compared to the results obtained from K-means clustering which used the multiple execution approach to determine the optimal number of clusters in each data set. In general, LNNAIS using SDOT tends to deliver clusters of similar or higher quality for all data sets, followed by LNNAIS using IPT and K-means clustering. The influence of the different LNNAIS parameters (using SDOT) was then investigated.

Chapter 7 presented and investigated different data migration types in a non-stationary environment. These migration types were pattern migration, cluster migration and centroid migration. A procedure to generate artificial non-stationary data sets with different environment parameters and migration types was proposed. Also, clustering performance measures for a non-stationary environment were proposed. The proposed clustering performance measures were used for comparison between four network based artificial immune system models for clustering of the generated artificial non-stationary data sets. A sensitivity analysis of the LNNAIS parameters shows that for all migration types, LNNAIS utilises small population sizes with small cluster sizes and larger population sizes for large cluster sizes. The clustering quality of LNNAIS is the lowest at high frequencies and high severities of change for all of the migration environments at different dimensions and cluster sizes. The clustering quality of LNNAIS improves with an increase in the cluster size at different dimensions. Increasing the number of dimensions lowers the clustering quality of LNNAIS at different cluster sizes. The clustering performance of the LNNAIS model and the enhanced version utilising SDOT (LNN_{SDOT}) were compared against the clustering performance of SMAIN and DWB in non-stationary environments. The higher quality of clusters found by SMAIN when compared to LNNAIS is due to a larger ALC population size which is utilised by SMAIN and overfits the data. This is more emphasised in cluster and centroid migration environments where the ALC population size of the SMAIN model does not scale with the number of clusters. LNNAIS delivers clusters of a higher quality than DWB and LNN_{SDOT} for all pattern and cluster migration environments. In centroid migration environments, LNNAIS succeeds to recover from any changes and improve on the cluster quality as time progresses, even though the number of clusters changes and clusters become less compact. LNNAIS delivers clusters of a higher quality than DWB for centroid migration environments and in cases where the data is not overfitted by SMAIN and the ALC population has similar sizes, LNNAIS also tends to deliver clusters of a higher quality than SMAIN. LNNAIS also obtains the correct number of clusters at different severities of change with no significant change in the ALC population size. Wherever LNN_{SDOT} obtained the correct number of clusters, the quality of the clusters tends to be higher than those clusters delivered by LNNAIS at different dimensions, clusters sizes, frequencies of change and severities of change. Furthermore, LNN_{SDOT} also tends to deliver clusters of higher quality than SMAIN and DWB in cases where the data is not overfitted by SMAIN and a similar number of clusters are obtained.

From the results presented in this thesis, it can be concluded that LNNAIS and LNN_{SDOT} are efficient clustering models for different stationary and non-stationary environments. This is achieved even in light of the smaller set of control parameters compared to other network based AIS models. LNN_{SDOT} can dynamically determine the number of clusters in a stationary data set and is most suitable for centroid migration non-stationary environments where the number of clusters in the data is not known and needs to be tracked over time.

8.2 Future Research

Several new directions for future research are briefly summarised below.

Decreasing neighbourhood sizes: Although the clustering performance of LNNAIS is the best at small neighbourhood sizes, a hybrid approach of a linear decrementing neighbourhood size needs to be investigated. An initial large neighbourhood size ($\rho = \mathcal{B}_{max}$) will have a more *greedy* approach to adapt the ALCs as one ALC network to the data patterns. More ALC networks are formed by linearly decreasing the neighbourhood size, which results into a more refined and specific search to the clusters in the data by different ALC networks. The model might initially prematurely adapt to the data, but eventually converge to different cluster centroids with the final set of ALC networks.

Alternative network neighbourhood topologies: The proposed LNNAIS in this thesis utilises a ring topology to determine the network connections between the ALCs. Although the clustering performance of LNNAIS at different neighbourhood sizes was investigated, future research needs to investigate the clustering performance of LNNAIS utilising different network topologies which includes rectangular grid (used in SOM), star and wheel (used in PSO) and Caylee trees. In addition to the investigation of the clustering performance of LNNAIS with different network topologies, the time of convergence and the coverage of the search space by the ALCs need to be investigated.

Hierarchical grouping: The clusters obtained by LNNAIS are represented by the formed ALC networks. The ALC networks are determined by pruning the network links between those ALCs with the lowest network affinity until the required number of clusters are obtained (or in the case of LNN_{SDOT} the number of clusters is dynamically determined by the outlier network affinities). There is a potential risk that at the time of pruning the network links to determine the ALC network boundaries, an ALC might have been in the process of adapting to a neighbouring ALC. This can result into an ALC which has a low network affinity between the ALC's predecessor and an even lower network affinity with the ALC's successor in the population. When the adapting ALC is grouped with an ALC network, the calculated mean of the ALCs in that network might not represent the most appropriate centroid of the cluster in the data, since the ALC is becoming an outlier to the network of ALCs. This will have an impact on the clustering performance of LNNAIS. Therefore, a hierarchical agglomerative approach needs to be investigated to determine whether there is less influence of adapting ALCs to the calculated centroid of an ALC network. Another potential risk is that the adapting ALC can have equal network affinities between its neighbouring ALCs. These network affinities might be the lowest in the population resulting in an ALC network which consists of a single ALC and which does not contain any data patterns. These risks of the behaviour of ALCs need to be investigated.

LNN$_{SDOT}$ with IPT: Since the sequential deviation outlier technique (SDOT) used by LNNAIS depends on outlier network affinities between the ALC networks in order to dynamically determine the ALC network boundaries, a hybrid approach of SDOT and the iterative pruning technique (IPT) needs to be investigated. SDOT can be used to determine the initial number of clusters as a starting value of K for IPT. IPT can then increment and/or decrement the value of K with each iteration. The stopping criteria depend on whether the validity index used by IPT is a monotonic increasing or decreasing function. In the case of a monotonic increasing function, if the validity index decreases with an increment or decrement in the value of K, the search terminates. In the case of a monotonic decreasing function, if the validity index increases with an increment or decrement in the value of K, the search terminates. The search continues in both cases until the stopping criteria are met. The hybrid approach will dynamically determine the number of clusters with SDOT if outlier network affinities exist, otherwise the IPT technique is initialised with the result of SDOT and the search continues with the IPT technique.

Generating non-stationary environments: The generated non-stationary environments in this thesis contained clusters with fixed and equal spreads. For all the migration types defined, further investigation is needed into the clustering performance of the models on non-stationary environments where the spread of clusters changes with migrating patterns. This means that with each migrated pattern joining a cluster, the spread of the cluster should increase by a certain ratio. The same reasoning should be followed for patterns migrating from a cluster. The spread of clusters from which patterns migrate should decrease by a certain ratio. The dynamic spread of clusters will result in non-stationary environments with clusters which not only have different sizes (as those used in this thesis), but also different spreads and densities.

Image segmentation and classification problems: The proposed LNNAIS can be applied to the problem of image segmentation and classification problems. Since LNNAIS is an unsupervised learning algorithm, no changes are necessary to apply LNNAIS to image segmentation problems. The pixels of an image are then seen as the data set of antigen patterns. The ALC population of LNNAIS will adapt to these antigen patterns by forming ALC networks and eventually cluster the pixels of the image. Each cluster represents a segment of the image. In the context of non-stationary environments, a sequence of images of specific scenery can be segmented to identify any moving objects in the image. Focusing on classification problems, LNNAIS needs to be changed in such a way that ALCs are labeled with the same class labels as in the antigen set of patterns. This means that ALCs can then only adapt to antigen patterns of the same class. Eventually each of the formed ALC networks will represent a specific class in the data set of antigen patterns. This is a more semi-supervised learning approach of LNNAIS for classification problems.