Dynamical Clustering via Neural Vector Fields with Attractor-Based Structure

Denys Symonov[1,*,†], Oleksandr Palagin[1,†] and Bohdan Zaika[1,†]

[1] V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine, Academician Glushkov Avenue, 40, Kyiv, 03187, Ukraine

Abstract

The paper considers the concept of data clustering based on a dynamic approach, in which clusters are defined not as geometric regions, but as attractors in a parameterised vector field. The model of Dynamic Attractor Clustering via Neural Vector Fields (DAC-NVF) is proposed, in which clustering is formalised as the evolution of objects according to a system of first-order differential equations. The vector field is modelled by a differentiated neural network and trained within the Neural ODE paradigm. Each object moves in space in accordance with the field, and its cluster affiliation is determined by the endpoint of the trajectory - the attractor to which it converges. The proposed model allows to automatically determine the number of clusters as the number of stable points of a dynamic system. The loss functionality includes three components: trajectory cohesion, attractor separation, and field smoothness regularisation. Objects with similar dynamics converge to a common attractor, forming a natural cluster. This approach enables the detection of complex topological data structure, particularly in cases where classical methods such as k-means or spectral clustering are insufficient due to their dependence on Euclidean metrics or a fixed number of clusters. To evaluate the effectiveness of the method, both standard quality metrics (Silhouette Score, Calinski-Harabasz Index, Davies-Bouldin Index) and specialised indicators developed for the dynamic model, such as Trajectory Stability Index (TSI) and Attractor Assignment Consistency (AAC), were used. The results of numerical modelling showed high stability and consistency of clustering. Comparison with classical methods has demonstrated the advantages of the DAC-NVF approach in problems with heterogeneous, high-dimensional data, in particular in the case of clustering profiles containing numerical and categorical features. Thus, the proposed method opens up new perspectives for clustering complex data by combining the mathematical rigour of dynamical systems models with the flexibility of neural network parameterisation.

Keywords

Dynamical clustering, Neural vector fields, Attractors, Neural ODE, Trajectory-based classification, Unsupervised learning, Topological structure of data, Trajectory Stability Index (TSI), Attractor Assignment Consistency (AAC)

1. Introduction

Clustering as a method of detecting latent structure in data sets is one of the fundamental tasks of modern mathematical modelling. In the broadest sense, it is a formalised way of dividing a set of objects into disjoint subsets (clusters), each of which unites objects that have internal similarities in certain respects. Over the past decades, clustering has become a central tool in a variety of fields: machine learning, bioinformatics, economics, social psychology, retail, and many other areas where segmentation of complex systems is important [1, 2].

Traditional approaches to clustering, such as k-means [1], hierarchical grouping algorithms [3], spectral methods [4], and Self-Organising Maps (SOM) [5], are based on assumptions that have

Workshop "Intelligent information technologies" UkrProg-IIT`2025 co-located with 15th International Scientific and Practical Programming Conference UkrPROG'2025, May 13-14, 2025, Kyiv, Ukraine

* Corresponding author.
† These authors contributed equally.
denys.symonov@gmail.com (D. Symonov); palagin_a@ukr.net (O. Palagin); zaikabohdan5@gmail.com (B. Zaika)
0000-0002-6648-4736 (D. Symonov); 0000-0003-3223-1391 (O. Palagin); 0009-0001-9567-8361 (B. Zaika)

both mathematical simplicity and serious limitations. Such assumptions include the Euclidean similarity metric, the isotropy of clusters, the need to fix their number in advance, and weak adaptability to complex, nonlinear, topologically heterogeneous structures in data. The application of these methods to real-world problems related to the analysis of multidimensional, mixed (categorical-numerical) or hierarchically organised data leads to the loss of important information and insensitivity to latent dynamics in the structure of the set.

In response to these challenges, differentiated approaches combining concepts from dynamical systems [6, 7], neural computing [8, 9], and multivariate analysis [10] have become widespread in recent years. One of these approaches is to formalise clustering as a problem of dynamic evolution of objects in a parameterised vector field, where clusters are modelled not as fixed geometric centres, but as stable fixed points - attractors - in the phase space. This approach involves not only the spatial proximity between objects, but also the analysis of their behaviour in time, which is determined by the dynamics of movement in an artificially constructed force field. The motion of objects towards attractors is described by a system of differential equations, and the field itself is formed and adapted in the course of training.

In this paper, we consider the problem of clustering objects represented by vectors of mixed types. In this paper, mixed-type vectors are multidimensional vectors that describe the complex nature of objects, such as customers in a marketing system [11], patients in medical data [12, 13], system specifications in engineering analysis [14], etc. These vectors often contain both numeric and categorical components, have a hierarchical structure, and thus cannot be adequately represented in a conventional Euclidean space without special pre-coding. In this context, the task of clustering objects represented by mixed-type vectors requires a new mathematical paradigm that allows taking into account not only metric but also dynamic similarity between objects.

The approach proposed in this paper, called Dynamic Attractor Clustering via Neural Vector Fields (DAC-NVF), is based on the idea that clusters should be viewed as regions of attraction in a dynamic field parameterised by a neural network. The dynamics of objects is modelled as trajectories in a multidimensional space that converge to stable points - attractors - in a force field. Thus, a cluster is not a static set, but the result of long-term system behaviour. The main advantages of this approach are that there is no need to fix the number of clusters in advance, adaptability to complex data geometry, and the ability to interpret the data structure in terms of dynamics and stability.

The vector field is trained based on a loss function that combines three components: cohesion (convergence to attractors), separation between attractors, and regularisation (smoothness of the field). The training is implemented within the framework of the Neural ODE concept, a model that provides automatic differentiation through an integrator, allowing accurate parameter updates based on the full trajectories of objects [15]. The attractors themselves are not predefined, but arise as a result of the system's evolution as its asymptotically stable points.

Dynamic clustering enables a natural modelling of complex data topology, including cases where clusters have a curved shape, variable density, or overlap. In addition, the presence of dynamics offers the possibility to study the behaviour of objects near cluster boundaries, which is impossible in classical approaches. It is also important that the system allows the construction of new, specialised metrics for assessing the quality of clustering, taking into account dynamic stability (e.g. Trajectory Stability Index) and clustering consistency during re-runs (Attractor Assignment Consistency).

The dynamic formulation of clustering provides a natural framework for modelling complex data topologies, particularly in cases where clusters exhibit non-convex shapes, varying densities, or partial overlaps. Moreover, the introduction of dynamics enables the examination of object behaviour near cluster boundaries—an aspect typically inaccessible to classical static approaches.

Significantly, the dynamic perspective also permits the development of novel, specialised clustering quality metrics that account for temporal or structural stability (e.g., the Trajectory Stability Index) and clustering consistency under repeated initialisations (e.g., Attractor Assignment Consistency).

The DAC-NVF method extends the scope of formal analysis in clustering by enabling not only the discovery of latent data organisation but also the interpretable assessment of cluster stability, topology, and sensitivity to initial conditions. Consequently, the proposed approach holds considerable potential for application in tasks where understanding the causal structure of data, forecasting based on dynamic evolution, and decision-making under trajectory-aware conditions are of critical importance.

Given the relevance and complexity of clustering under conditions of high dimensionality, dynamism, and mixed-type features, this work addresses a mathematically significant problem: the construction of a clustering model that is not only algorithmically efficient, but also conceptually aligned with the principles of modern dynamical systems theory, flow-based geometry, and neural representational learning.

2. Overview of existing approaches and limitations

Clustering, that is, the partitioning of a set of objects into homogeneous subsets, is one of the fundamental problems in mathematical modelling, statistical learning, and applied data analysis. In the context of high-dimensional spatial structures—particularly those arising from the analysis of hyperparameter profiles—clustering serves not only as a tool for uncovering latent organisation, but also as a means of formalising internal dynamics within systems characterised by complex feature structures.

Today, there are a number of well-known clustering methods that can be divided into several conceptual classes: metric-oriented (e.g., k-means [16]), topological (Self-Organising Maps [5]), spectral (Spectral Clustering [4]), and graph (Community Detection [17]). This section discusses the most common methods of cluster analysis, their theoretical foundations, as well as limitations from the standpoint of modern problems of classifying objects with a complex structure.

The most classical example is the k-means method, which is aimed at minimising the intra-cluster variance for a given number of clusters *K*. Formally, the following problem is solved:

$$\min_{\{C_k\}_{k=1}^{K}} \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2, \tag{1}$$

where μ_k is the centre of the cluster C_k; $\|\cdot\|$ is the Euclidean norm.

The k-means method is simple to implement, computationally efficient, and scales well on large samples. However, its analytical model has a number of limitations:

- the hypothesis that the clusters are isotropic and close to convex Euclidean regions;
- the need to pre-determine *K*;
- the sensitivity to initialisation and the presence of local minima.

These assumptions make k-means unsuitable for objects with nonlinear, non-Euclidean structure or complex topology.

Self-Organising Maps implement a neural network approach to clustering by training a grid of neurons, where each neuron has an associated weight vector. The training is based on bringing the input vectors closer to the nearest neurons according to a metric criterion and gradually adjusting the grid weights while preserving the local topology. The advantages of the SOM method include the following:

- preservation of the topological structure of the data;
- ability to perform non-linear clustering;
- interpretability through visual maps (U-matrix).

However, SOM has certain limitations:

- the need for a predefined lattice geometry;
- the inability to adaptively determine the number of clusters;
- the lack of clear boundaries between clusters;
- non-differentiality, which makes it difficult to integrate with deep learning (since SOM does not support backpropagation of gradients).

Spectral clustering methods use the spectrum (eigenvalues and vectors) of a Laplacian matrix built from the metrics between objects. The basic idea is to project the data into a space where the cluster structure becomes linearly separable, and then apply k-means. Although spectral methods are capable of detecting nonlinear boundaries between clusters, they have a number of analytical and computational limitations:

- the need to compute the spectrum of the Laplacian matrix, which has a computational complexity of $O(N^3)$;
- sensitivity to the choice of kernel hyperparameters and the number of neighbours in the graph;
- lack of model differentiability;
- poor scalability for large object sets.

Thus, spectral clustering has high theoretical power but limited practical applicability without specific optimisations.

Despite the diversity of approaches, most classical clustering methods share common limitations:

- metric dependency: they rely on static distance measures, without accounting for the dynamics or evolution of objects;
- rigid structure: they require pre-specifying the number of clusters, assume convex cluster shapes, and depend on the symmetry of the metric space;
- lack of integrated learning dynamics: they are incapable of continuous learning within a differentiable framework;
- instability to perturbations and local changes: they exhibit weak invariance of cluster structure under data modification;
- inability to represent heterogeneous object descriptions: traditional methods fail to account for the presence of categorical, hierarchical, or vectorial features within an object's structure.

In particular, in the case of clustering mixed-type vectors that combine numerical and categorical characteristics, have a complex internal dependence and a high level of multidimensionality, none of the traditional approaches is able to provide simultaneously:

1. automatic determination of the number of clusters;
2. stable topological classification;
3. mathematical consistency of the dynamics of cluster formation;
4. possibility of further differentiated optimisation.

In light of the above, there is a clear need to develop models that:

- interpret clustering not as a geometric partitioning, but as a dynamic process of object evolution within a given field;
- enable the formation of cluster structures as a consequence of dynamic behaviour, rather than the result of local criterion optimisation;

- ensure asymptotic stability of clusters through the formation of attractors;
- support the use of neural network parameterisation and automatic differentiation;
- are applicable to objects represented as mixed-type vectors, which lack a straightforward Euclidean representation.

Such approaches must integrate the mathematical theory of differential dynamical systems, Lyapunov stability theory, neural network-based approximation analysis, as well as modern deep learning techniques, to build systems capable of adaptive, stable, and topologically consistent clustering. The following sections will demonstrate that such a model is feasible—specifically, in the form of clustering via neuronally parameterised vector fields, where clusters emerge as asymptotically stable regions (attractors) of the dynamical system.

3. The goal of the research

The goal of this researh is to build, theoretically substantiate and practically implement a method of clustering hyperparametric objects based on dynamics in a neural-parametric vector field, where clusters are formed as attractors. In particular, the tasks are as follows:

- to formulate a mathematical model of dynamics and cluster formation;
- to build a generalised loss functional that takes into account the cohesion, stability and separation of clusters;
- to develop an algorithm for learning the parameters θ using numerical integration and automatic differentiation methods;
- to analyse the stability of attractors and justify the automatic determination of the number of clusters as a consequence of the structure of the vector field.

Thus, the study aims to integrate the theoretical foundations of dynamical systems with practical machine learning methods to build a stable and interpretable clustering of hyperparametric objects.

4. Problem statement

Clustering is a fundamental task in data analysis, especially in cases where each object is characterised by a complex structure - a set of interdependent variables, which we will further treat as mixed-type parameters. In this context, mixed-type parameters are descriptive characteristics of an object that may belong to a certain cluster. An example of an object is a client profile represented by a vector of numerical and categorical attributes (age, gender, activity, region, etc.).

Let's assume that $\mathcal{H} \subseteq \mathbb{R}^p \times C_1 \times \ldots \times C_m$ is the space of all possible vectors of mixed-type parameters, $\mathcal{H} = \{h^{(i)}\}, i = \overline{1,n}$, where $h^{(i)}$ is the representation of an individual or entity; $\mathbb{R}^p$ is the space of numerical characteristics; C_j is the finite set of possible values of the *j*-th categorical parameter, $j = \overline{1,m}$.

In order to work in the numerical space, a vectorisation function $\phi: \mathcal{H} \to \mathbb{R}^n$ is introduced, which transforms all vectors of mixed-type parameters into a space of constant dimension $\mathbb{R}^n$ (for example, through one-hot, embedding, or ordinal encoding).

The task of clustering is to build a cluster structure of the set $\{x_i\}, i = \overline{1,n}$, where $x_i = \phi(h^{(i)}) \in \mathbb{R}^n$, which:

- does not require a pre-fixed number of clusters;
- reflects the topological and dynamic proximity between objects;
- enables to consider clusters as pulling areas (attractors) in a certain force field.

The clustering of objects in a vectorised space makes it possible to interpret clusters as attractors that reflect the dynamic organisation of objects in a multidimensional space.

5. Modelling of dynamics and attractors

This paper proposes an approach to clustering based on the concept of dynamic convergence of objects to attractors in a controlled force field represented by a vector function. Clustering is considered not as a measurement of Euclidean proximity to centres, but as a dynamic system where each object moves towards stable points in space - attractors that arise as a result of neural network training.

The proposed approach is to define a parameterised vector field $F_\theta: \mathbb{R}^n \to \mathbb{R}^n$, which defines the dynamics of object movement in the space of mixed-type parameters. Each object $x_i \in \mathbb{R}^n$ is considered as the initial point of a trajectory that develops according to the following equation:

$$\frac{dx_i(t)}{dt} = F_\theta(x_i(t)), \tag{2}$$

where $x_i(0) = x_i$, $\theta \in \mathbb{R}^m$ are parameters that define the field structure (for example, weights and shifts in the neural network layers).

This system defines the trajectory of the object in time, which develops according to the field structure F_θ. The function F_θ is trained in such a way that all trajectories converge to a limited number of asymptotically stable stationary points (attractors) $\xi_k \in \mathbb{R}^n$.

A stationary point ξ_k is said to be an attractor if the condition $F_\theta(\xi_k) = 0$ is satisfied, and the linearised system in the neighbourhood of ξ_k, defined by the Jacobian matrix:

$$J(\xi_k) = \nabla F_\theta(\xi_k), \tag{3}$$

has a spectrum which real parts are strictly negative (Lyapunov stability):

$$\mathrm{Re}\left(\lambda_i\big(J(\xi_k)\big)\right) < 0, \forall i. \tag{4}$$

In this case, ξ_k is an asymptotically stable fixed point, and there exists a neighbourhood $U_\xi \subset \mathbb{R}^n$ such that $x_i(0) \in U_\xi \Rightarrow \lim_{t\to\infty} x_i(t) = \xi_k$. The set of all such initial points defines the attractor circle ξ_k:

$$\mathcal{B}(\xi_k) = \left\{x_i \in \mathbb{R}^n: \lim_{t\to\infty} \Phi_T^\theta(x_i) = \xi_k\right\}. \tag{5}$$

In the context of clustering, each region $\mathcal{B}(\xi_k)$ is interpreted as a cluster, and ξ_k as its central representative. The clusters produced by this model satisfy the following properties:

- deterministic clustering: each object follows a unique trajectory that converges to a single attractor ξ_k;
- automatic determination of the number of clusters: the number of attractors k is not predefined but emerges during training as the number of isolated stable points of the field F_θ;
- cluster stability: small perturbations in x_i do not affect cluster membership, provided the object remains within $\mathcal{B}(\xi_k)$;
- topological coherence: objects within the same cluster exhibit similar dynamic behaviour.

In contrast to classical methods that rely on Euclidean or spectral metrics, the proposed approach operates on the geometry of trajectories in the space. The distance between objects is not a primary criterion for clustering; rather, it is their behaviour within the field F_θ, defined by the learned vector field, that determines the structure. This enables the discovery of non-linear cluster formations that are difficult to detect using traditional methods.

6. Loss function

The field F_θ is trained by minimising the loss function, which is defined as:

$$\frac{dx_i(t)}{dt} = F_\theta(x_i(t)), \tag{6}$$

where $\mathcal{L}_{conv}$ is the convergence function that minimises the intra-cluster variance; $\mathcal{L}_{separ}$ is the cluster separation component; $\mathcal{R}_{regular}$ is the penalty for the complexity or non-smoothness of the field; $\lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$ are the weighting coefficients.

Let $x_i(T)$ be the vector position of the mixed-type parameters after numerical integration in the field F_θ up to time *T*:

$$x_i(T) = \Phi_T^\theta(x_i), \tag{7}$$

under the condition (2).

Then the convergence function has the following form:

$$\mathcal{L}_{conv} = \sum_{i=1}^{N} \min_k \|x_i(T) - \xi_k\|^2, \tag{8}$$

where $\{\xi_k\}$ are approximate attractor points defined as stable fixed points.

To avoid excessive merging of attractors, a cluster separation component $\mathcal{L}_{separ}$ is added, which moves attractors away from each other to preserve the distinctiveness of the clusters:

$$\mathcal{L}_{separ} = \sum_{1 \leq k \leq l \leq K} \frac{1}{\|\xi_k - \xi_l\|^2 + \delta}, \tag{9}$$

where $\delta > 0$ is a small term to avoid degeneracy.

To ensure smoothness and manageability of the dynamics, the following regularisation is used:

$$\mathcal{R}_{regular} = \int_\Omega \|\nabla F_\theta(x_i)\|^2 \, dx, \tag{10}$$

or an empirical approximation on a discrete grid:

$$\mathcal{R}_{regular} \approx \frac{1}{M} \sum_{j=1}^{M} \left\|\nabla F_\theta(z_j)\right\|^2, \tag{11}$$

where $\{z_j\}$ are random or uniformly taken points.

In summary, the overall loss function $\mathcal{L}(\theta)$ takes into account both the convergence of the trajectories of mixed-type parameters with the nearest attractors ($\mathcal{L}_{conv}$) and the structural separation between clusters ($\mathcal{L}_{separ}$), supplemented by a regulariser ($\mathcal{R}_{regular}$) that controls the complexity and smoothness of the vector field F_θ. This combined formulation provides a balanced field training aimed at forming stable, separated and dynamically consistent cluster structures in the space of mixed-type parameters.

7. Learning algorithm

The process of training a parameterised vector field F_θ, which implements clustering in the form of a dynamic system, is based on the numerical integration of object trajectories in the field and optimisation of the loss functional using the gradient method. For each object $h^{(i)} \in \mathcal{H}$, after its transformation $\phi(h^{(i)}) = x_i \in \mathbb{R}^n$, the dynamic system is integrated according to condition (2), until a fixed time *T*, or until the stationary state $\|F_\theta(x_i(t))\| < \varepsilon$ is reached. Numerical integration is performed by the Runge-Kutta method or an adaptive integrator within the Neural Ordinary Differential Equations (Neural ODE) model [15].

To accurately and efficiently train the parameters θ, we use the Neural ODE concept. This approach involves differentiation through an integrator, allowing us to train the parameters of a vector field based on the full trajectory.

The vector field is implemented as a small fully connected neural network, where:

- input: $x_i \in \mathbb{R}^n$, $F_\theta(x_i) \in \mathbb{R}^n$;
- structure: input layer (n neurons), hidden layers (2-4 layers of 64-128 neurons each), activation function (tanh or softplus), output layer (n neurons).

The Neural ODE allows you to calculate the derivatives of the loss in terms of θ without having to explicitly store all the intermediate integration values. This is implemented through the adjoint sensitivity method, which calculates:

$$\frac{d\mathcal{L}(\theta)}{d\theta} = -\int_0^T \alpha(t)^T \frac{\partial F_\theta(x_i)}{\partial \theta} dt, \tag{12}$$

where $\alpha(t) = \frac{\partial \mathcal{L}(\theta)}{\partial (x_i)}$ is the sensitivity vector.

Upon completion of the evolution of each object, the endpoint $x_i(T)$ is obtained, which is interpreted as a potential realisation of the attractor ξ_k. In order to build a cluster structure, the points $x_i(T)$ are aggregated into subsets according to the proximity criterion, for example, based on the threshold radius δ. Each such subset determines the empirical attractor $\hat{\xi}_k$, and the corresponding objects - the cluster membership. At each learning iteration, the loss function is computed, which includes the computation of $\mathcal{L}_{conv}$, $\mathcal{L}_{separ}$ and $\mathcal{R}_{regular}$.

The resulting gradient of the loss function with respect to the parameters θ determines the direction of the neural network update:

$$\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}(\theta), \tag{13}$$

where η is the learning rate.

The iterative learning process enables the adaptive formation of the vector field F_θ, in which the dynamics of objects converges to a limited number of stable attractors ξ_k, forming a cluster structure $C_k := \mathcal{B}(\xi_k)$, under condition (5).

8. Comparison of clustering methods

Table 1 provides a summary of the advantages and limitations of various approaches, serving as a conceptual comparison between the proposed method and three classical clustering algorithms. The comparison is based on key criteria: the requirement to predefine the number of clusters, the ability to capture complex geometric structures in data, model differentiability, support for topological cohesion, scalability to large datasets, and the incorporation of internal dynamics within the model.

Given that the objects in the study are vectors of mixed-type parameters, classical methods do not take into account the complex internal structure of such vectors (the presence of numerical and categorical components, hierarchical dependence, etc.). Instead, we propose a model:

- works with any type of parameters after vectorisation;
- models the interaction between their components through the field structure;
- provides clustering that is robust, topologically meaningful, and dynamically controlled.

Table 2 shows the results of the statistical analysis of the test data set.

Table 1
Comparing the properties of clustering methods

Criterion	DAC-NVF	k-means	SOM	Spectral Clustering
The need to fix *K*	No	Yes	Yes	Yes
Support of complex geometry	Yes	No	Partially	Yes
Differentiability	Yes	No	No	No
Topological cohesion	Yes	No	Yes	Partially
Scalability	High	High	High	Low
Presence of force dynamics	Yes	No	No	No

Table 2
Statistical characteristics of the test dataset

Metrics	Age	AnnualIncome	SpendingScore	MembershipYears	SatisfactionLevel
Count	25000	25000	25000	25000	25000
Mean	39.7	52158.69	49.51	5.01	0.286
Standard Dev.	11.54	16965.57	24.02	2.25	0.166
Minimum	18	10000	1	0	0
25^{th} percentile	31	40636.04	32	3	0.16
50^{th} percentile	40	52028.71	49	5	0.27
75^{th} percentile	48	63594.44	67	6	0.39
Maximum	70	123145.46	100	7	1

Based on Table 2, the test dataset contains information on 25,000 customers and is characterised by high variability of parameters, which allows for a detailed analysis of the behavioural and socio-economic characteristics of the customer base. The age distribution is close to normal with an average of 40 years, and the annual income fluctuates in a wide range, having an average of about 52,000 units with a significant dispersion. The SpendingScore and SatisfactionLevel also show a wide range of values, indicating a heterogeneity of customer habits and attitudes. The data is suitable for cluster analysis, building predictive models, identifying behavioural patterns, and developing personalised customer interaction strategies.

Table 3 shows a comparison of the main clustering quality metrics for the four approaches, which allows us to assess their strengths and weaknesses in the context of data analysis.

Table 3
Comparison of the main clustering quality metrics

Quality metric	DAC-NVF	k-means	SOM	Spectral Clustering
Silhouette Score	0.097	0.11	0.15	0.17
Calinski-Harabasz Index	589	611	307	208
Davies-Bouldin Index	2.29	2.29	1.82	1.71

The analysis of comparative clustering metrics (Table 3) shows that none of the approaches is dominant by all criteria, but each has its advantages depending on the data structure. The dynamic method showed the highest intra-cluster cohesion (CH Index ≈ 589), but at the same time, poor geometric cluster separation (Silhouette = 0.097), which is explained by its focus on dynamic rather than static topology. The k-means method showed similar results, indicating a linear structure of clusters in the data. The expected improvement for SOM confirms its potential with proper tuning, and spectral clustering showed the ability to extract clusters with less overlap, albeit with reduced inter-cluster differentiation. Overall, the dynamic approach reveals a deeper topological structure that is not fully captured by standard metrics, which highlights the feasibility of using additional dynamic indicators.

Low values of geometric metrics do not mean that the quality of dynamic clustering is poor - they only indicate that this method requires different evaluation approaches. Let's analyse two criteria for this purpose. The first criterion, the Trajectory Stability Index (TSI), evaluates how stable the trajectories of objects in a dynamic field are under a small perturbation of the initial conditions and is defined as follows:

$$TSI = \frac{1}{N}\sum_{i=1}^{N} \exp\left(-\frac{\left\|x_i(T) - x_i^{\delta}(T)\right\|^2}{\epsilon^2}\right), \tag{14}$$

where ϵ is the tolerance scale.

The second criterion, Attractor Assignment Consistency (AAC), assesses whether the same attractors will be assigned to objects after the field is re-run (possibly with a random initialisation) and is defined as follows:

$$AAC = \frac{1}{N}\sum_{i=1}^{N} \left(\mathrm{Mode}\left(\xi_k^{(1)}, \xi_k^{(2)}, \dots, \xi_k^{(n)}\right) = \xi_k^{(1)}\right). \tag{15}$$

The results of calculating specialised metrics for the dynamic approach to clustering (DAC-NVF) of the test sample confirm its structural stability. The value of $TSI = 0.739$ indicates high local stability of trajectories: most objects show a convolution to the same attractors even under small perturbations of the initial conditions. The metric $AAC = 0.927$ indicates the reliability of cluster formation: more than 92% of objects received the same cluster assignment when the system was re-run. This demonstrates that the model is not only consistent, but also capable of reproducible clustering in a mixed-type parameter space, which is crucial for analysing complex systems with a mixed nature of features.

Figure 1 shows the results of data clustering in projection on the first two principal components (PCA) for the four methods: DAC-NVF (dynamic approach), k-means, SOM and Spectral Clustering.

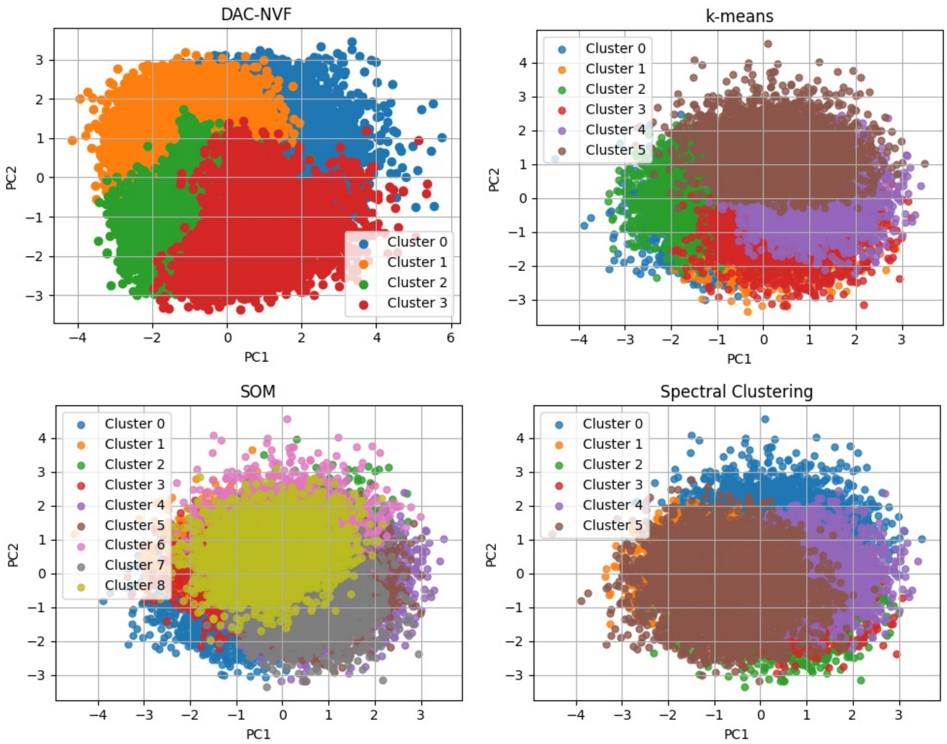

Figure 1: Visual results of cluster formation for the test dataset.

Visually, it is clear that DAC-NVF forms well-separated clusters that occupy meaningful areas of space, which confirms its ability to model attractive regions. The k-means method demonstrates partial overlap of clusters with radial placement around the centre, typical for Euclidean optimisation. In the case of SOM, an excessive number of clusters with strong overlap is observed, indicating the need for additional lattice tuning. Spectral Clustering also produces large areas of cluster overlap, although some areas have clear boundaries, which is a consequence of building on local similarity. In general, DAC-NVF reveals a better topological structure, while the other methods show geometric but less stable clustering.

9. Conclusions and directions for further research

In this paper, we formulate and implement a mathematical paradigm of clustering based on the idea of modelling the cluster structure as a dynamic behaviour of objects in a neurally parameterised vector field. In contrast to traditional methods based on distance metrics, isotropic hypotheses and Euclidean geometry, the proposed approach considers clusters as areas of attraction (attractors) in phase space where the trajectories of objects converge under the influence of a vector field defined by a system of differential equations.

The key achievement of the developed model is its ability to adaptively detect the number of clusters, form stable structures in a multidimensional space, and provide differential training of field parameters based on the Neural ODE concept. The model demonstrates the ability to reproduce complex topologies, including irregular, non-Euclidean and variable-density clusters, which is unattainable for most classical algorithms. At the same time, the process of cluster formation does not occur through direct grouping by distance, but as a consequence of the asymptotic behaviour of the system, thus maintaining the correspondence between the field structure and the cluster organisation of the set.

The results of experiments on test data have shown that the dynamic clustering model is able to provide high topological cohesion, invariance to initial perturbations, and consistency of the cluster structure during re-runs. This is confirmed by the values of dynamic metrics: Trajectory Stability Index ($TSI = 0.739$) and Attractor Assignment Consistency ($AAC = 0.927$), which capture both the local stability of trajectories and the global consistency of clustering results. It is also worth highlighting the high value of the Calinski-Harabasz Index (≈ 589), which indicates deep intra-cluster cohesion, despite the fact that geometric metrics such as Silhouette Score do not fully reflect the advantages of the dynamic approach.

From a methodological point of view, the presented approach is important for building integrated clustering models, where the cluster structure is the result of not only geometric but also functional and dynamic organisation of objects. Within the proposed paradigm, it is possible to move from passive data analysis to active modelling of the processes that determine the structure of a set. In this way, clustering appears not as a tool for dividing data, but as a means of studying the latent dynamics of complex systems, which is directly related to the tasks of forecasting, control and optimisation.

The applicability of the developed method to the clustering of objects described by vectors with heterogeneous elements (numerical, categorical, textual, structural) deserves special attention. Such vectors are typical for modern tasks in marketing, healthcare, behavioural analytics, and engineering systems. The proposed model provides not only clustering of such objects, but also their interpretation through the properties of attraction, which opens up prospects for explanatory machine learning.

Several important areas for further research are worth highlighting:

- Theoretical analysis of the attractor set: building conditions for the existence, isolation, and stability of the attractor set depending on the field parameters and the distribution of objects.

- Generalisation to stochastic dynamical systems: extending the model by introducing stochastic perturbations and analysing stability in the Ito or Lyapunov sense in the mean.
- Extension to semi-supervised learning: integration of partial labels or domain information into the field structure as external sources of control over cluster formation dynamics.
- Optimisation of computational procedures: acceleration of trajectory integration and vector field learning through the use of adaptive integrators, low-rank approximations and graph structures.
- Analysis of multi-layer vector fields: construction of deep hierarchies of dynamics, where attractors of one level are initialisations on the next, which allows modelling a multi-level cluster structure.

Thus, the presented approach has both direct applied significance - for cluster analysis tasks in real systems - and a deep theoretical perspective - as a new class of models that combine dynamical systems, topology, and deep learning. Its development can contribute to the formation of a new research area in mathematical clustering - dynamically oriented clustering, where the concepts of attraction, stability and evolution of objects become fundamental in modelling the structure of a set.

Acknowledgements

The work was supported by the state budget research project "Develop methods for modelling the processes of targeted management of complex multi-component information systems for various purposes" (state registration number 0123U100754) of the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL in order to translate research notes and results from Ukrainian to English. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] T. Dinh, W. Hauchi, P. Fournier-Viger, D. Lisik, M.-Q. Ha, H. C. Dam, V.-N. Huynh, Categorical data clustering: 25 years beyond K-modes, Expert Systems with Applications 272 (2024) 126608. doi:10.48550/arXiv.2408.17244.

[2] C. Liang, S. Wang, M.-J. Zhang, T. Li, Diagnosis, clustering, and immune cell infiltration analysis of m6A-related genes in patients with acute myocardial infarction—a bioinformatics analysis, Journal of Thoracic Disease 14 (2022) 1607–1619. doi:10.21037/jtd-22-569.

[3] Z. Pang, C. Wang, L.-F. Zhao, Y. Liu, G. Sharma, Cross-Modality Hierarchical Clustering and Refinement for Unsupervised Visible-Infrared Person Re-Identification, IEEE Transactions on Circuits and Systems for Video Technology 34 (2024) 2706–2718. doi:10.1109/TCSVT.2023.3310015.

[4] J. Xie, W. Kong, S. Xia, G. Wang, X. Gao, An Efficient Spectral Clustering Algorithm Based on Granular-Ball, IEEE Transactions on Knowledge and Data Engineering 35 (2023) 9743–9753. doi:10.1109/TKDE.2023.3249475.

[5] S. Mueller, D. Kranzlmueller, Self-Organising Maps for Efficient Data Reduction and Visual Optimisation of Stereoscopic based Disparity Maps, CSRN (2022). doi:10.24132/csrn.3201.28.

[6] L. Caprini, D. Breoni, A. Ldov, et al., Dynamical clustering and wetting phenomena in inertial active matter, Communications Physics 7 (2024) 343. doi:10.1038/s42005-024-01835-y.

[7] M. Becchi, F. Fantolino, G. M. Pavan, Layer-by-layer unsupervised clustering of statistically relevant fluctuations in noisy time-series data of complex dynamical systems, Proceedings of

the National Academy of Sciences of the United States of America 121 (2024). doi:10.1073/pnas.2403771121.
[8] E. Mello Rella, A. Chhatkuli, E. Konukoglu, et al., Neural Vector Fields for Implicit Surface Representation and Inference, International Journal of Computer Vision 133 (2025) 1855–1878. doi:10.1007/s11263-024-02251-z.
[9] X. Yang, G. Lin, Z. Chen, L. Zhou, Neural Vector Fields: Implicit Representation by Explicit Learning, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 16727–16738. doi:10.1109/CVPR52729.2023.01605.
[10] V. Khare, R. S. Shukla, S. Pandey, S. K. Singh, Identification of Drought Tolerant Recombinant Inbred Lines and Estimation of Genetic Diversity by Clustering and Multivariate Analysis in Wheat RILs Populations, International Journal of Bio-resource and Stress Management (2022). doi:10.23910/1.2022.2924.
[11] L. Blasco-Arcas, H.-H. M. Lee, M. N. Kastanakis, M. Alcañiz, A. Reyes-Menendez, The role of consumer data in marketing: A research agenda, Journal of Business Research 146 (2022) 436–452. doi:10.1016/j.jbusres.2022.03.054.
[12] T. H. Vo, G. Chauvet, A. Happe, E. Oger, S. Paquelet, V. Garès, Extending the Fellegi-Sunter record linkage model for mixed-type data with application to the French national health data system, Computational Statistics & Data Analysis 179 (2023) 107656.
[13] O. Palagin, D. Symonov, T. Semykopna, Mathematical Modeling of the Evolution of the Rehabilitation Process for Patients with Oncological Diseases, in: V. Kazymyr et al. (Eds.), Mathematical Modeling and Simulation of Systems. MODS 2023, Lecture Notes in Networks and Systems, vol. 1091, Springer, Cham, 2024, pp. 99–112. doi:10.1007/978-3-031-67348-1_8.
[14] T. Wilschut, A. T. Hofkamp, T. J. L. Schuijbroek, L. Etman, J. Rooda, Bridging the gap between requirements engineering and systems architecting: the Elephant Specification Language, Design Science (2024). doi:10.1017/dsj.2024.30.
[15] I. Okubo, K. Sugiura, H. Matsutani, A Cost-Efficient FPGA-Based CNN-Transformer Using Neural ODE, IEEE Access 12 (2024) 155773–155788. doi:10.1109/ACCESS.2024.3480977.
[16] Z. Huang, H. Zheng, C. Li, C. Che, Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection, Academic Journal of Science and Technology (2024). doi:10.54097/74414c90.
[17] D. B. Acharya, H. Zhang, Community Detection Clustering via Gumbel Softmax, SN Computer Science 1 (2020). doi:10.1007/s42979-020-00264-2.