

# Spatio-Temporal Clustering Benchmark for Collective Animal Behavior

Eren Cakmak, Manuel Plank, Daniel S. Calovi, Alex Jordan, Daniel Keim

Centre for the Advanced Study of Collective Behaviour  
University of Konstanz, Germany

## ABSTRACT

Various spatio-temporal clustering methods have been proposed to detect groups of jointly moving objects in space and time. However, such spatio-temporal clustering methods are rarely compared against each other to evaluate their performance in discovering moving clusters. Hence, in this work, we present a spatio-temporal clustering benchmark for the field of collective animal behavior. Our reproducible benchmark proposes synthetic datasets with ground truth and scalable implementations of spatio-temporal clustering methods. The benchmark reveals that temporal extensions of standard clustering algorithms are inherently useful for the scalable detection of moving clusters in collective animal behavior.

## CCS CONCEPTS

- Computing methodologies → Cluster analysis; • General and reference → Experimentation; • Information systems → Clustering.

## KEYWORDS

Spatio-Temporal Clustering, Trajectory Clustering, Benchmark, Moving Clusters, Collective Animal Behavior.

### ACM Reference Format:

Eren Cakmak, Manuel Plank, Daniel S. Calovi, Alex Jordan, Daniel Keim. 2021. Spatio-Temporal Clustering Benchmark for Collective Animal Behavior. In *1st ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility (HANIMOB'21), November 2, 2021, Beijing, China*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3486637.3489487>

## 1 INTRODUCTION

Spatio-temporal clustering is crucial for analyzing groups of moving objects in various applications, such as in transportation analysis or the study of collective animal behavior. A central task in such domains is to identify moving clusters, a group of objects moving close together in space and time [9]. However, identifying such moving clusters remains challenging due to often large-scale datasets, resulting in a trade-off between computational cost and accuracy. In addition, the performance of spatio-temporal clustering methods

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*HANIMOB'21, November 2, 2021, Beijing, China*

© 2021 Association for Computing Machinery.  
ACM ISBN 978-1-4503-9122-1/21/11...\$15.00  
<https://doi.org/10.1145/3486637.3489487>

is rarely compared against each other, posing a challenge for users to select accurate and scalable methods.

Hence, in this work, we present a benchmark of spatio-temporal clustering in the field of collective animal behavior. The benchmark proposes 3600 datasets with varying data characteristics to compare the performance of different common spatio-temporal clustering methods. We believe that our benchmark enhances the experimental reproducibility of spatio-temporal clustering results within animal movement ecology. For the benchmark datasets and implemented methods, please refer to our online project page.<sup>1</sup> In summary, our main contributions are: (1) A diverse set of synthetic collective behavior datasets with ground-truth, (2) a reproducible benchmark of spatio-temporal clustering algorithms, and (3) scalable implementations of spatio-temporal clustering methods.

## 2 BACKGROUND

Research in collective animal behavior is focused on the interactions and emergent behaviors of animal groups [12]. Much of this research is framed from the perspective of self-organization, exploring how global patterns emerge from local rules of interaction. Therefore, a core aspect of understanding animal collectives is understanding the spatio-temporal rules of interactions governing group behavior. This general framework has numerous use cases, from social insects [5], to shoals and schools of fish [20], and groups of primates [18]. From a biological perspective, it is essential to know whether individuals form loose and temporary associations, which promotes only weak recognition and social complexity, from groups that form long-term, stable associations with familiar individuals [14]. Thus, understanding the nature and frequency of interactions in animal collectives represents a core goal of the field. Analyzing movement patterns helps to discover collectives, such as the clustering of flocks [11], swarms [15], or convoys [22]. Here, we use the term moving clusters as described by Dodge et al. [9] instead of domain-specific terms, such as flocks, swarms, or convoys.

Spatio-temporal clustering methods can be classified into techniques for trajectory and moving object clustering [15]. Trajectory clustering usually uses specific geometric distance metrics (e.g., dynamic time warping or Hausdorff distance) to compute similarities between the mover's trajectories and utilizes afterward standard clustering techniques (e.g., K-means) [21]. Moving object clustering discovery, on the other hand, identifies similar movement behavior directly by adapting classical clustering algorithms to spatio-temporal data, such as the spatio-temporal extension of DBSCAN [4]. For further reading on spatio-temporal clustering, we refer to the surveys of Kisilevich et al. [13], Yuan et al. [21], Atluri et al. [3], and Ansari et al. [2].

---

<sup>1</sup>[github.com/eren-ck/spatio-temporal-clustering-benchmark](https://github.com/eren-ck/spatio-temporal-clustering-benchmark)

In contrast to previous work, we present a spatio-temporal clustering benchmark comparing methods against each other to evaluate their performance. Until now, experimental studies usually compare spatio-temporal clustering methods against baselines on custom datasets. For instance, Agrawal et al. [1] propose ST-OPTICS and compare the method against ST-DBSCAN [4]. As there is no unified and commonly used experimental dataset and protocol, it remains challenging to compare the performance of spatio-temporal clustering methods. Therefore, we propose a benchmark for detecting moving clusters in collective animal behavior to overcome these prevailing shortcomings, focusing on generated synthetic data with ground truth, and presenting state-of-the-art baseline methods.

### 3 BENCHMARK DESIGN

#### 3.1 Problem Statement

Spatio-temporal clustering aims to detect jointly moving objects in space and time without having any labels. Intuitively a moving cluster can be seen as a sequence of static spatial clusters with the objects being spatially close to each other during the whole sequence. Identifying such moving clusters is valuable for various applications in animal movement ecology, such as analyzing migrating bird flocks. In such applications, we cannot cluster the spatio-temporal data with standard clustering methods (e.g., DBSCAN [10]) due to the temporal dimension. Hence, detecting such moving clusters requires adopting clustering methods utilizing similarity metrics that partition both the spatial and temporal data dimensions. Ideally, such spatio-temporal clustering methods are scalable to large-scale datasets, handle high-dimensional data with additional attributes, and discover arbitrary cluster shapes [1].

#### 3.2 Dataset Design

The goal of our dataset design was to generate realistic spatio-temporal datasets with ground truth. We used three collective behavior models to generate synthetic datasets with known ground truth clusters, covering all the main features of existing models.

First, the seminal Reynolds model [17] is based on fixed speed agents which change the direction of their movement according to three basic concepts: if two agents are too close together, they will repel each other, each agent has a contribution to move towards the center of its neighbors inside an interaction radius and to align their movement with these same neighbors. It is fundamental to notice that from the three described models, the Reynolds model is the only one without inertia, meaning an agent is allowed to invert its direction of movement in one time step completely.

Second, the Couzin model [8], similarly to the Reynolds model, has the same three behaviors, but each in a different zone, which depends on the distance of the focal agent. If an agent has one or more neighbors at a distance  $R$  smaller than  $R_{zor}$ , it will only present repulsion in the opposite direction. If a focal agent has no neighbors to which it is repelled, it will try to match the orientation of all neighbours that are at a distance  $R$  where  $R_{zor} \geq R < R_{zoo}$ , and are attracted to all neighbours that are in the zone  $R_{zoo} \geq R < R_{zoa}$ . Unlike the Reynolds model, here we have a limitation to a maximum turning rate  $\theta$  of how fast an agent can reorient itself, and a blind angle  $\alpha$  directly behind the direction of movement, where the focal agent ignores the contribution of neighbors located there.

Third, the Gautrais-Calovi model is a data-driven model based on the experiments by Gautrais and collaborators [6]. Movement is described as a persistent turning walker model (PTW), meaning that, differently from the previous models, all interactions act on the turning speed of the agent instead of directly affecting its heading. In this model, agents do not have a defined radius of interaction or an abrupt regime change depending on the distance: agents interact with the agents comprising the first shell of the Voronoi neighbors. Each agent in this first shell will always contribute to the attraction and alignment of the focal agent. The contribution is modulated by a continuous function that goes to zero directly ahead and behind the agent, as an analogous (but continuous) blind zone as seen in the Couzin model. The model is the only one that does not have a repulsion contribution since, in their original experiments, the authors observed that agents used depth to avoid collisions.

Some of the principal results coming from these models are phase transitions from order and disorder (seen in all three models) and the existence of different behavioral models, such as polarized (schooling) and vortex (milling) for the Couzin and Gautrais-Calovi models. The Gautrais-Calovi model has also investigated in detail these behavioral modes and observed the existence of clear intermittence between the polarized and vortex state, where the group will spontaneously shift from one state to the other.

We used data generation models several times with different parameters and later concatenated them to obtain datasets with clusters. We created 30 datasets with a length of 50,000 time steps, varying numbers of movers, clusters, and different input parameters for each model to obtain datasets with distinct characteristics, e.g., different cluster densities. We enclosed the exact input parameter configurations for each dataset with our datasets.

Afterward, we randomly sampled from the produced 90 datasets a variety of different sized datasets, with a varying number of moving objects, clusters, and timesteps. We likewise added noise by randomly sampling and attaching single movers from the initial 90 datasets. Furthermore, we also subsampled moving objects from the original moving clusters randomly over time, for instance, we randomly subsampled 20 movers from a cluster with 50 movers. Through the sampling, we generated diverse datasets with different temporal lengths, cluster densities, uncertainty, and noise. We generated 100 datasets for the three collective behavior models with the temporal lengths of 100, 300, 600, 900, 1200, 1500, 1800, 2100, 2400, 2700, 3000, and 3500. Every dataset has up to ten clusters with up to 20 agents per cluster. The dataset sizes range from 1200 up to 520,000 data points. As a result, we obtained 3600 spatio-temporal datasets with a diverse set of data characteristics.

#### 3.3 Evaluation Metrics

We compared the performance and execution time of the spatio-temporal clustering methods. We measured the clustering quality using the ground truth with the adjusted mutual information (AMI) score [19]. We employed the adjusted metric to ensure that uniform random label assignments will result in scores close to zero. In addition, we captured the run-time in seconds of each algorithm for all synthetic datasets. Moreover, we limited the clustering run-time to 120 seconds, meaning that if an algorithm takes longer than two minutes, we canceled the respective clustering method.

## 4 EXPERIMENTS

### 4.1 Methods

We compare two standard baseline methods against a set of temporal extensions of standard clustering algorithms. The two baselines are ST-DBSCAN [4] and trajectory clustering using Euclidean distance as a similarity measure.<sup>2</sup>

Our implemented clustering algorithms extend standard algorithms in two ways by (1) processing large-scale datasets by splitting the datasets into frames and thus (2) generalizing them to discover clusters in spatio-temporal clusters. We argue that we can adjust most clustering algorithms to accommodate not only spatial but also temporal features. Two approaches inspired our temporal extensions of clustering algorithms. First, ST-DBSCAN [4] uses two distance parameters to assess whether a data point is density-reachable from another data point. Second, the splitting and merging process for spatio-temporal data by Peca et al. [16]. Building on these ideas, we (1) split the dataset periodically into overlapping frames, (2) filter the distance matrix of data points using a temporal distance parameter, (3) employ standard clustering algorithms, and afterward (4) merge the results of subsequent frames. Only subsets of the original data are loaded into memory by splitting the temporal dimension into frames using a fixed time window. Two subsequent frames also always have an overlapping period to ensure that we can merge the resulting cluster labels of individual frames. We benchmark the outlined extension for the following scalable clustering methods: DBSCAN [10], agglomerative clustering, K-Means, BIRCH, and HDBSCAN [7] using the following implementations.<sup>3</sup> Moreover, we implemented spatio-temporal extensions of spectral and affinity propagation clustering methods. However, even for smaller datasets, both methods required more than two minutes of execution time.

We perform a simplistic hyperparameter search by grid search for the 3600 datasets and each clustering method. The searched parameter space examines suitable parameter combinations, which we specified based on the generated data characteristics. For instance, we determine the number of clusters for the temporal extension of K-Means using the dataset ground truth. To guarantee the successful merging of cluster labels across frames, we choose an overlap of 10% between two subsequent frames. In summary, we compare two baseline approaches against a set of scalable temporal extensions of standard clustering algorithms. We evaluate the clustering quality and execution time in detecting moving clusters.

### 4.2 Experimental Setup

All experiments were computed on a server with 30 CPU cores (Intel Xeon CPU E5-2640 v3 @ 2.60GHz) and 162 GB RAM. The setup with so much main memory is necessary for the baseline methods. For example, ST-DBSCAN calculates a distance matrix between all points and thus has a quadratic memory consumption. We excluded the parameter search in our run-time analysis. If the underlying clustering implementations were parallelized, we employed the parallelization functionality.

<sup>2</sup>Using implementations of: [github.com/GISerWang/Spatio-temporal-Clustering](https://github.com/GISerWang/Spatio-temporal-Clustering) and [github.com/cshjarry/trajecotry\\_cluster](https://github.com/cshjarry/trajecotry_cluster).

<sup>3</sup>Using implementations of: [github.com/scikit-learn/scikit-learn](https://github.com/scikit-learn/scikit-learn) and [github.com/scikit-learn-contrib/hdbscan](https://github.com/scikit-learn-contrib/hdbscan).

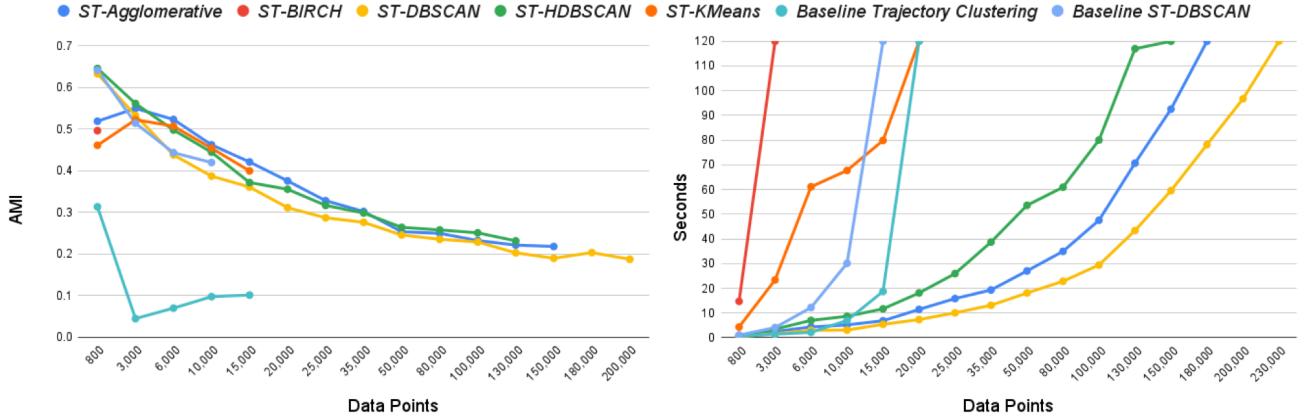
### 4.3 Results

Fig. 1 shows the respective results for grouped dataset sizes. For instance, the first group 800 encompasses all data sets containing 800-3000 data points. The AMI results show that the temporal extensions of the standard algorithms and the ST-DBSCAN baseline method achieve comparable clustering quality for small datasets, for instance, less than 20,000 data points). However, the baseline ST-DBSCAN method does not scale to large datasets due to the quadratic memory consumption. Moreover, the trajectory clustering method using the Euclidean distance fails to detect moving clusters, even for smaller datasets. In addition, we identified three scalable clustering methods, specifically the temporal extensions of HDBSCAN, DBSCAN, and agglomerative clustering. For larger datasets, the AMI decreases due to increasing cluster overlaps and consequently growing merging errors for subsequent overlapping frames. Our temporal extension of ST-DBSCAN scales the furthest, up to 200,000 data points. The execution time highlights the performance of each method. Some methods only scale to small data sets, such as the baseline ST-DBSCAN or the baseline trajectory clustering method. The standard clustering method's temporal extensions are scaling better considering the datasets are split into subsequent frames and then merged again, requiring no quadratic memory consumption. However, the potential merging process also causes defects due to possible spatial overlaps within the overlapping frames. Furthermore, none of the evaluated methods detected moving clusters in datasets larger than 200,000 points within two minutes of run-time.

The benchmark results show the performance of our temporal extensions of standard clustering methods is useful to detect moving clusters in collective animal behavior datasets. We recommend using our temporal extension of ST-DBSCAN or ST-HDBSCAN to identify moving clusters if the number of clusters is not known in advance. If the number of clusters is available, we recommend agglomerative clustering. Overall, the proposed methods enable analyzing group dynamics of swarms, flocks, and other animal collectives. We believe that the implemented spatio-temporal clustering methods are crucial to detect group and sub-groups of moving clusters in collective animal behavior. For example, one can use the proposed methods to study evolving structures within collectives by exploring temporally stable and changing sub-groups.

### 4.4 Limitations

First, we mainly focused on the generation and comparison of 2D spatio-temporal datasets. However, our implemented clustering methods can also be used to cluster and benchmark 3D spatio-temporal datasets, such as the 3D movement of a fish shoal. Moreover, the clustering methods have several parameters that need to be specified. With a simple parameter search, we tried to set the hyperparameters for our benchmark. However, such parameters have to be set manually by the user in real-world applications. For example, the frame size strongly influences the run-time and the clustering quality, leading to the computation of larger distance matrices and possibly resulting in fewer merging errors. We also decided to use a run-time constraint of two minutes to limit the execution time of the methods. Likewise, our datasets partitioning into groups (e.g., 800-3000 data points) resulted in differing dataset



**Figure 1: The results of our benchmark. The average adjusted mutual information (AMI) score and the average run-time for different dataset group sizes are displayed. For example, the group 800 includes all datasets containing 800–3000 data points.**

numbers in each group due to random data generation and sampling. In our benchmark, we did not investigate moving clusters with a varying number of movers over time. As a result, movers are not able to switch from one cluster to another. Moreover, the synthetic datasets are not modeling specific animal species but rather capture general movement patterns of collectives.

## 5 CONCLUSION

We presented an initial step towards practical algorithms for finding spatio-temporal clusters in collective animal behavior. We generated numerous synthetic datasets and compared the performance of different spatio-temporal clustering algorithms. Based on our benchmark, we report that temporal extensions of standard clustering methods (e.g., ST-HDBSCAN) are valuable and scalable methods to detect moving clusters in the field of collective animal behavior.

## ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984.

## REFERENCES

- [1] KP Agrawal, Sanjay Garg, Shashikant Sharma, and Pinkal Patel. 2016. Development and validation of OPTICS based spatio-temporal clustering technique. *Information Sciences* 369 (2016), 388–401.
- [2] Mohd Yousuf Ansari, Amir Ahmad, Shehzroz S Khan, Gopal Bhushan, et al. 2019. Spatiotemporal clustering: a review. *Artificial Intelligence Review* 53, 4 (2019), 1–43.
- [3] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–41.
- [4] Derya Birant and Alp Kut. 2007. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering* 60, 1 (2007), 208–221.
- [5] Eric Bonabeau, Guy Theraulaz, Jean-Louis Deneubourg, Serge Aron, and Scott Camazine. 1997. Self-organization in social insects. *Trends in ecology & evolution* 12, 5 (1997), 188–193.
- [6] Daniel S Calovi, Ugo Lopez, Sandrine Ngo, Clément Sire, Hugues Chaté, and Guy Theraulaz. 2014. Swarming, schooling, milling: phase diagram of a data-driven fish school model. *New Journal of Physics* 16, 1 (jan 2014), 015026. <https://doi.org/10.1088/1367-2630/16/1/015026>
- [7] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, Berlin, Heidelberg, 160–172.
- [8] Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. 2002. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* 218, 1 (2002), 1–11.
- [9] Somayeh Dodge, Robert Weibel, and Anna-Katharina Lautenschütz. 2008. Towards a taxonomy of movement patterns. *Information Visualization* 7, 3-4 (2008), 240–252.
- [10] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (Portland, Oregon) (KDD'96). AAAI Press, 226–231.
- [11] Joachim Gudmundsson and Marc Van Kreveld. 2006. Computing longest duration flocks in trajectory data. In *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems*. Association for Computing Machinery, New York, NY, USA, 35–42.
- [12] Andrew J King, Gaëlle Fehlmann, Dora Biro, Ashley J Ward, and Ines Fürtbauer. 2018. Re-wilding collective behaviour: an ecological perspective. *Trends in Ecology & Evolution* 33, 5 (2018), 347–357.
- [13] Slava Kisilevich, Florian Mansmann, Mirco Nanni, and Salvatore Rinzivillo. 2010. Spatio-temporal clustering. In *Data Mining and Knowledge Discovery Handbook*. Springer, Boston, MA, 855–874.
- [14] Etienne Lein and Alex Jordan. 2021. Studying the evolution of social behaviour in one of Darwin's Dreamponds: a case for the Lamprologine shell-dwelling cichlids. *Hydrobiologia* 848, 16 (2021), 1–28.
- [15] Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays. 2010. Swarm: Mining Relaxed Temporal Moving Object Clusters. *Proceedings of the VLDB Endowment* 3, 1 (2010), 723–734.
- [16] Iulian Peca, Georg Fuchs, Katerina Vrotsou, Natalia V Andrienko, and Gennady L Andrienko. 2012. Scalable Cluster Analysis of Spatial Events. *EuroVA@EuroVis* 6 (2012), 19–23.
- [17] Craig W Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. Association for Computing Machinery, New York, NY, USA, 25–34.
- [18] Ariana Strandburg-Peshkin, Damien R Farine, Iain D Couzin, and Margaret C Crofoot. 2015. Shared decision-making drives collective movement in wild baboons. *Science* 348, 6241 (2015), 1358–1361.
- [19] Nguyen Xuan Vinh, Julien Epps, and James Bailey. 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research* 11 (2010), 2837–2854.
- [20] Ashley JW Ward, David JT Sumpter, Iain D Couzin, Paul JB Hart, and Jens Krause. 2008. Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Sciences* 105, 19 (2008), 6948–6953.
- [21] Guan Yuan, Penghui Sun, Jie Zhao, Daxing Li, and Canwei Wang. 2017. A review of moving object trajectory clustering algorithms. *Artificial Intelligence Review* 47, 1 (2017), 123–144.
- [22] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. 2008. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 312–321.