Trajectory Grouping Structure: the Video

Kevin Buchin*
Department of Mathematics
and Computer Science
TU Eindhoven
k.a.buchin@tue.nl

Maike Buchin
Faculty of Mathematics
Ruhr-University Bochum
Maike.Buchin@rub.de

Marc van Kreveld
Department of Information and
Computing Sciences
Utrecht University
m.j.vankreveld@uu.nl

Bettina Speckmann*
Department of Mathematics
and Computer Science
TU Eindhoven
b.speckmann@tue.nl

Frank Staals*
Department of Information and
Computing Sciences
Utrecht University
f.staals@uu.nl

Introduction. In recent years there has been an increase in location-aware devices and wireless communication networks. This has led to a large amount of trajectory data capturing the movement of animals, vehicles, and people. The increase in trajectory data goes hand in hand with an increasing demand for techniques and tools to analyze them, for example, in transportation sciences, sports, ecology, and social services.

An important task is the analysis of movement patterns. In particular, given a set of moving entities we wish to determine when and which subsets of entities travel together. When a sufficiently large set of entities travels together for a sufficiently long time, we call such a set a *group*. Groups may start, end, split and merge with other groups. Apart from the question what the current groups are, we also want to know which splits and merges led to the current groups, when they happened, and which groups they involved. We wish to capture this group change information in a model that we call the *trajectory grouping structure*.

The Video. This note accompanies a video which explains the trajectory grouping structure and showcases results of our implementation using simulated NetLogo flocking data [3, 4] and real-world data from the Starkey project [2]. Below we briefly describe the trajectory grouping structure and sketch some of our results. Additional details can be found in the (full) paper [1].

A Definition for a Group. The informal definition above suggests that three parameters are needed to define groups: (i) a spatial parameter for the distance between entities; (ii) a temporal parameter for the duration of a group; (iii) a count for the number of entities in a group. We use the three parameters as follows: a small spatial parameter implies we are interested only in spatially close groups, a large temporal parameter implies we are interested only in long-lasting groups, and a large count implies we are interested only in

large groups. By adjusting the parameters suitably, we can obtain more detailed or more generalized views of the grouping structure.

More formally, let $\mathcal{X}$ be a set of entities of which we have locations during some time span. The *ε-disc* of an entity x (at time t) is a disc of radius ε centered at x at time t. Two entities are *directly connected* at time t if their ε-discs overlap. Two entities x and y are *ε-connected* at time t if there is a sequence $x = x_0, .., x_k = y$ of entities such that for all i, x_i and x_{i+1} are directly connected.

A subset $S \subseteq \mathcal{X}$ of entities is ε-connected at time t if all entities in S are pairwise ε-connected at time t. This means that the union of the ε-discs of entities in S forms a single connected region. The set S forms a *component* at time t if and only if S is ε-connected, and S is maximal with respect to this property. The set of components $\mathcal{C}(t)$ at time t forms a partition of the entities in $\mathcal{X}$ at time t.

Let the spatial parameter of a group be ε, the temporal parameter δ, and the size parameter m. A set G of k entities forms a *group* during time interval I if and only if the following three conditions hold: (*i*) G contains at least m entities, so $k \geq m$, (*ii*) the interval I has length at least δ, and (*iii*) at all times $t \in I$, there is a component $C \in \mathcal{C}(t)$ such that $G \subseteq C$.

We denote the interval $I = [t_s, t_e]$ of group G with I_G. Group H *covers* group G if $G \subseteq H$ and $I_G \subseteq I_H$. If there are no groups that cover G, we say G is *maximal* (on I_G). In Fig. 1, groups $\{x_1, x_2\}$, $\tilde{G} = \{x_3, x_4\}$, $\hat{G} = \{x_5, x_6\}$, and $G = \{x_1, .., x_4\}$ are maximal: $\tilde{G}$ and $\hat{G}$ on $[t_0, t_5]$, G on $[t_1, t_2]$. Group $\{x_1, x_3\}$ is covered by G and hence not maximal.

Note that entities can be in multiple maximal groups at the same time. For example, entities $\{y_1, y_2, y_3\}$ can travel together for a while, then y_4, y_5 may become ε-connected, and shortly thereafter y_1, y_4, y_5 separate and travel together for a while. Then y_1 may be in two otherwise disjoint maximal groups for a short time. An entity can also be in two maximal groups where one is a subset of the other. In that case the group with fewer entities must last longer. That an entity is in more groups simultaneously may seem counter-
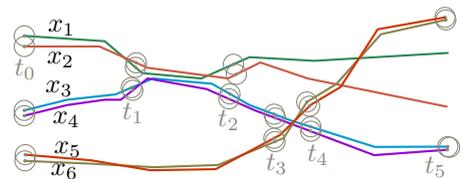
Figure 1: For $m = 2$ and $\delta > t_4 - t_3$ **there are four maximal groups:** $\{x_1, x_2\}$, $\{x_3, x_4\}$, $\{x_5, x_6\}$, **and** $\{x_1, .., x_4\}$.

intuitive, but it is necessary to capture all grouping information.

Our maximal group definition uses three parameters, which all allow a more global view of the grouping structure. In particular, we observe that there is *monotonicity* in the group size and the duration: If G is a group during interval I, and we decrease the minimum required group size m or decrease the minimum required duration δ, then G is still a group on time interval I. Also, if G is a maximal group on I, then it is also a maximal group for a smaller m or smaller δ. For the spatial parameter ε we observe monotonicity in a slightly different manner: if G is a group for a given ε, then for a larger value of ε there exists a group $G' \supseteq G$. The monotonicity property is important when we want to have a more detailed view of the data: we do not lose maximal groups in a more detailed view. The group may, however, be extended in size and/or duration.

We capture the grouping structure using a *Reeb* graph of the ε-connected components together with the set of all maximal groups. Parts of the Reeb graph that do not support a maximal group can be omitted. The grouping structure can help us in answering various questions. For example:

- What is the largest/longest maximal group at time t?
- How many entities are currently (not) in any maximal group?
- What is the first maximal group that starts/ends after time t?
- What is the total time that an entity was part of any maximal group?
- Which entity has shared maximal groups with the most other entities?

The grouping structure can also be used to partition the trajectories in independent data sets, to visualize grouping aspects of the trajectories, and to compare grouping across different data sets.

Representing the Grouping Structure. Let $\mathcal{X}$ be a set of n entities, where each entity travels along a path of τ edges. To compute the grouping structure we consider a manifold $\mathcal{M}$ in $\mathbb{R}^3$, where the z-axis corresponds to time. The manifold $\mathcal{M}$ is the union of n "tubes". Each tube consists of τ skewed cylinders with horizontal radius ε that we obtain by tracing the ε-disc of an entity x over its trajectory. Let H_t denote the horizontal plane at height t, then the set $\mathcal{M} \cap H_t$ is the *level set* of t. The connected components in the level set of t correspond to the components (maximal sets of ε-connected entities) at time t.

The Reeb Graph. We start out with a possibly disconnected solid that is the union of a collection of tube-like regions: a 3-manifold with boundary. Note that this manifold is not explicitly defined. We are interested in horizontal cross-sections, and the evolution of the connected components of these cross-sections. To describe how the components change over time, we consider the Reeb graph $\mathcal{R}$ of $\mathcal{M}$. The Reeb graph has a vertex v at every time t_v where the components change. The vertex times are usually not at any of the given times $t_0, .., t_\tau$, but in between two consecutive time steps. A directed edge $e = (u, v)$ connecting vertices u and v, with $t_u < t_v$, corresponds to a set C_e of entities that form a component at any time $t \in I_e = [t_u, t_v]$. The Reeb graph is this directed graph. Note that the Reeb graph depends on the spatial parameter ε, but not on the other two parameters of maximal groups.

THEOREM 1. *Given a set $\mathcal{X}$ of n entities, in which each entity travels along a trajectory of τ edges, the Reeb graph $\mathcal{R} = (V, E)$ has $O(\tau n^2)$ vertices and edges. This bound is tight in the worst case.*

The Trajectory Grouping Structure. The trajectories of entities are associated with the edges of the Reeb graph in a natural way.
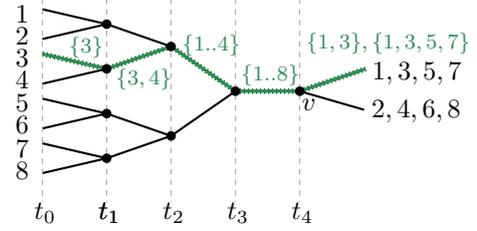

Figure 2: The maximal groups containing entity 3 **(green). Vertex** v **creates six new groups, including** $\{1, 3\}$ **and** $\{1, 3, 5, 7\}$**.**

Each entity follows a directed path in the Reeb graph from a start vertex to an end vertex. Similarly, (maximal) groups follow a directed path from a start or merge vertex to a split or end vertex. If $m > 0$ or $\delta > 0$, there may be edges in the Reeb graph with which no group is associated. These edges do not contribute to the grouping structure, so we can discard them. The remainder of the Reeb graph we call the *reduced Reeb graph*, which, together with all maximal groups associated with its edges, forms the *trajectory grouping structure*.

Bounding the Number of Maximal Groups. To bound the total number of maximal groups, we study the case where $m = 1$ and $\delta = 0$, because larger values can only reduce the number of maximal groups. It may seem as if each vertex in the Reeb graph simply creates as many maximal groups as it has outgoing edges. However, consider for example Fig. 2. Split vertex v creates not only the maximal groups $\{1, 3, 5, 7\}$ and $\{2, 4, 6, 8\}$, but also $\{1, 3\}$, $\{5, 7\}$, $\{2, 4\}$, and $\{6, 8\}$. These last four groups are all maximal on $[t_2, t]$, for $t > t_4$. Notice that all six newly discovered groups start strictly before t_v, but only at t_v do we realize that these groups are maximal, which is the meaning that should be understood with "creating maximal groups". This example can be extended to arbitrary size. Hence a vertex v may create many new maximal groups, some of which start before t_v.

THEOREM 2. *Let $\mathcal{X}$ be a set of n entities, in which each entity travels along a trajectory of τ edges. There are at most $O(\tau n^3)$ maximal groups, and this is tight in the worst case.*

Computing the Grouping Structure. We first sweep through time to compute the Reeb graph in $O(\tau n^2 \log n)$ time. While computing the Reeb graph we already identify which entities "live" on which edges of the graph. We then compute the maximal groups with a second sweep through time.

THEOREM 3. *Given a set $\mathcal{X}$ of n entities, in which each entity travels along a trajectory of τ edges, we can compute all maximal groups in $O(\tau n^3 + N)$ time, where N is the output size.*

REFERENCES

[1] K. Buchin, M. Buchin, M. van Kreveld, B. Speckmann, and F. Staals. Trajectory grouping structure. In *Proc. 13th Algorithms and Data Structures Symposium (WADS)*, LNCS 8037, pages 219–230, 2013. **arXiv:1303.6127 [cs.CG]**

[2] Oregon Department of Fish and Wildlife and the USDA Forest Service. The Starkey project, 2004.

[3] U. Wilensky. NetLogo flocking model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1998.

[4] U. Wilensky. NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999.