

On Measures for Groups of Trajectories

Lionov Wiratma, Marc van Kreveld and Maarten Löffler

Abstract We present a list of measures for a single trajectory, including measures that require the presence of other trajectories, such as the centrality of a trajectory amidst other trajectories. Then, we introduce three different views in order to extend measures of a single trajectory to a group, namely the representative view, the complete view and the area view. Furthermore, we give measures that exist only for a group of trajectories, like density and formation stability. We also show that it may be possible to define new measures by combining trajectory data with data from other sources, such as the environment where the entities move. Finally, we discuss several tasks: settlement selection, visualization and segmentation, where measures on groups of trajectories are necessary.

Keywords Trajectories · Groups of trajectories · Movement attributes · Measures

1 Introduction

With the increased use and quality of GPS and other positioning devices, the analysis of *trajectory* data has become a mainstream topic in GIScience. A trajectory is the model of a moving entity. Trajectory data comes, for example, from vehicle, animal, hurricane, pedestrian, and sports player tracking.

We can use the *abstract model* or the *data model* for trajectories. In the abstract model, we consider the trajectory to be the representation of a moving object,

L. Wiratma (✉) · M. van Kreveld · M. Löffler
Department of Information and Computing Sciences, Utrecht University,
Utrecht, The Netherlands
e-mail: l.wiratma@uu.nl; lionov@unpar.ac.id

M. van Kreveld
e-mail: m.j.vankreveld@uu.nl

M. Löffler
e-mail: m.loffler@uu.nl

L. Wiratma
Department of Informatics, Parahyangan Catholic University, Bandung, Indonesia

assumed to be a point, that moves without discontinuities. In the data model, we realize that data on a moving object is generally collected by a device that gives a location at certain times where the location is sampled.

In the abstract model, a *trajectory* is a function T that maps a time interval $I = [t_\alpha : t_\beta]$ to the plane or 3-space. The time interval is the domain of this function. The image of the function on I is referred to as the *path* of the trajectory. The path has a shape and a direction but no time component. The location at the start time t_α is the *origin* of T and the location at the end time t_β is the *destination* of T .

In the (GPS) data model, a trajectory is a time-ordered sequence of triples $(x_1, y_1, t_1), \dots, (x_m, y_m, t_m)$ where at time t_i , the moving object was recorded to be at location (x_i, y_i) . We essentially do not know where the object was at times in between the samples. We could assume that locations are (sufficiently) precise, or incorporate imprecision in the analysis. The sampling rate and geometric precision are the two most important aspects of data quality for trajectories. Note that in the abstract model such quality aspects do not exist. If the sampling rate is sufficiently high and the precision as well, we can interpolate location and time between consecutive triples to make a continuous mapping from time to space. The triples can still be the most suitable representation.

We can distinguish the following main types of trajectory analysis:

- Segmentation (Anagnostopoulos et al. 2006; Buchin et al. 2011c): Partitioning a trajectory into segments so that within each segment, certain attributes of the trajectory are uniform. It may be used to split a trajectory into different movement behaviors.
- Similarity analysis (Buchin et al. 2011b; Liu and Schneider 2012): Analysing how much two trajectories appear alike. It can be used to determine how similar movement of a single entity is on different days, or to determine how similar the movement of two different entities is. The reasons for appearing alike can be rather varied: we could define similarity based on visiting the same locations, or based on having the same speed development (e.g. slow in the beginning, then linearly increasing speed).
- Clustering (Nanni and Pedreschi 2006; Lee et al. 2007): Grouping the trajectories in a collection based on similarity.
- Outlier detection (Lee et al. 2008a): Finding trajectories that do not belong to any cluster.
- Classification (Lee et al. 2008b): Assigning a trajectory to a cluster of trajectories, based on similarity.
- Hotspot detection (Gudmundsson et al. 2013): Finding places that are visited frequently and/or by many trajectories.
- Pattern detection: Finding interesting characteristics in one or more trajectories on any sort. Examples are leadership patterns (Andersson et al. 2007) and commuting patterns (Buchin et al. 2011a).
- Flocks, group, herd detection (Benkert et al. 2008; Buchin et al. 2015; Huang et al. 2008): Finding a special type of pattern determined by spatial proximity of a subset of the entities over a period of time. It is related to clustering, but

in clustering we generally consider the whole trajectory when making clusters, whereas in grouping a single entity can be in different groups at different times, or even at the same time.

All analysis types depend on *measures* defined on trajectories. In the listing above, we mentioned location, speed, and similarity as attributes that are defined in the abstract model, and can be computed in the data model. In fact, an attribute like speed gives rise to multiple measures: average speed, variation of speed, total stationary time, etcetera.

The purpose of this paper is to discuss measures for *groups of trajectories*. Recently, a number of different definitions of groups have been given, accompanied by algorithms that compute them from a collection (Hwang et al. 2005; Jeung et al. 2008; Buchin et al. 2015). Many analysis tasks that apply to trajectories exist for groups of trajectories as well. For example, we can imagine segmentation of the whole group at once, or doing a similarity analysis on two groups. To perform such analyses, we also need measures for the whole group. Groups are a natural unit of aggregation which can be analysed at once and which can be visualized using a single stroke. With the ongoing trend of dealing with larger and larger collections of data, aggregation is one of the approaches to cope with the growth.

In this paper we first provide an overview of the measures that exist for trajectories. We classify them into three types: measures for a single trajectory in isolation, measures for a single trajectory amidst other trajectories, and measures for a single trajectory in other environments. Then we proceed with extensions of these measures to groups, while at the same time including new measures that do not exist for single trajectories. We use the same classification into three types. When developing group measures, similar approaches are sometimes taken. We will highlight such more general approaches because they can potentially be used for other measures needed in specific applications.

2 Measures for a Single Trajectory

When defining measures for trajectories we will use the abstract model, since it is mathematically more clean and also representation-independent. All measures can be converted in various ways to measures in a data model.

Since a trajectory is a function from a time interval to the plane (or to space), the image of the function gives the locations where the entity is at the relevant times. The derivative of the function is the velocity, which is a vector, and the second derivative is acceleration, which is also a vector. Velocity has two components, namely speed and heading. These at any time measurable features (location, speed, heading, ...) of a trajectory are called *attributes*, and attributes are often the basis of measures defined over a whole trajectory. For example, from speed we can derive the measures *average speed*, *maximum speed*, *standard deviation of speed*, *percentage of standstill*, and more. These measures give a single value for the whole trajectory.

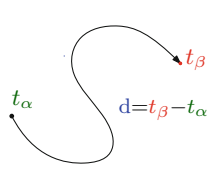
There are also attributes for trajectories amidst other trajectories. For example, for a given trajectory we have the attribute closest distance to another trajectory, which exists at any time. We can use this as the basis for a measure that intuitively corresponds to *degree of isolation* by averaging.

In this section we give an overview of various general-purpose measures that can be defined for a single trajectory. We first discuss measures for a single trajectory in isolation, then measures for a single trajectory amidst other trajectories, and then measures for a single trajectory in various contexts.

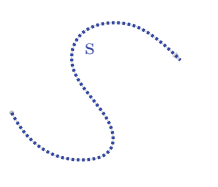
2.1 Measures for a Trajectory in Isolation

In this section we describe measures that relate to a single trajectory independent of other trajectories and the environment. Our list is not complete; we describe a number of existing useful, general-purpose measures. We begin with a number of basic, well-known measures capturing properties of the whole trajectory at once. Note that most of these common measures also appeared in similar lists of measures given in previous research (Laube et al. 2007; Dodge et al. 2008; Andrienko et al. 2008, 2013).

Duration

Description	<i>The length of the time interval of the trajectory (d)</i>			
Unit	second	Range	$[0, \infty)$	
Derived from	-			
Related works	(Calenge et al. 2009)			

Traversed distance

Description	<i>The total length of the path of the trajectory (s)</i>			
Unit	meter	Range	$[0, \infty)$	
Derived from	location			
Related works	(Yuan et al. 2010)			

Average speed

Description	<i>The ratio of the traversed distance and the duration (v)</i>			
Unit	meter per second	Range		$[0, \infty)$
Derived from	speed			
Related works	(Buard et al. 2011)			

While the average speed is a useful measure, it tells little about the variation in speed during the time interval. The entity could have been moving with constant speed equal to the average speed, or it could have stood still for half the time and then moved to the destination with speed double the average. Since this distinction is often important, we would like to have measures for it. The most obvious candidate is the standard deviation of speed over I .

In principle speed is an attribute that has a value at any time in I . To capture speed development over I in a more extensive manner, there are many options, both quantitative and qualitative. For example, we could split up I into k equal-duration subintervals and use the average speed in each of them. This may not be a good description of speed development because it might be more fine-grained than a chosen value of k . Ideally, one wants a partition of I into subintervals where the speed is similarly behaved. This can be obtained by segmentation.

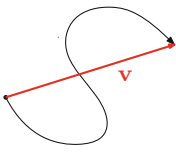
A derivative of speed is *acceleration* (and deceleration). Since acceleration exists at any time, we have a similar measure choices as for speed. For example, we can use the standard deviation of the acceleration to describe its variation.

Global direction

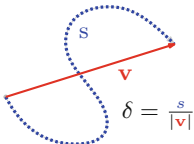
Description	<i>The direction of the vector from the origin to the destination of T (θ)</i>			
Unit	radian	Range		$[0, 2\pi]$
Derived from	location			
Related works	(LaPoint et al. 2013; Safi et al. 2013; Ranacher and Tzavella 2014)			

Also direction can be described more finely than with just a global direction, and as with speed, we can use subintervals of I to describe it. These subintervals may be obtained by equal durations or by segmentation.

Global velocity

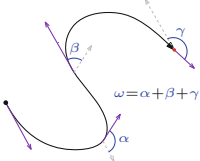
Description	<i>The vector from the origin to the destination of T divided by total duration (\mathbf{v})</i>		
Unit	vector	Range \mathbb{R}^2	
Derived from	velocity, or location and time		
Related works	(Hanks et al. 2011)		

Detour

Description	<i>The ratio of the traversed distance and the length of the global velocity vector (δ)</i>		
Unit	-	Range $[1, \infty)$	
Derived from	-		
Related works	(Buchin et al. 2011c; Boinski and Garber 2000)		

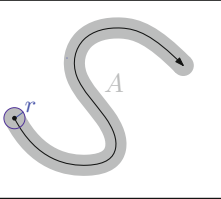
In other works, the notion of detour is described under different names, for example, sinuosity (Andrienko et al. 2013) and straightness index (Benhamou 2004).

Total angular change

Description	<i>The total change of the heading angle (ω)</i>		
Unit	radian	Range $[0, \infty)$	
Derived from	heading		
Related works	(Calenge et al. 2009)		

The *total angular change* is a common shape descriptor that does not depend on time. It is the global version of the attribute representing the change in heading. It describes the shape in just one number on an angular scale.

Area covered

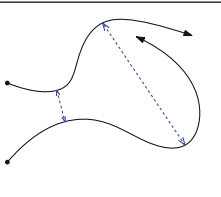
Description	<i>The area covered by a disc with radius r that moves along the trajectory (A)</i>		
Unit	square meter (m^2)	Range $ [0, \infty)$	
Derived from	location		
Related works	(Giuggioli et al. 2011)		

Finally, it may be useful to define the area covered by the trajectory. Andrienko et al. (2013) describe this measure as the *spatial extent* of a trajectory, which can be defined in several ways. A simple definition is to use a distance parameter r and say that the moving entity covers the whole area within distance r from its location. The covered area is then the total swept area of a disk centered at the entity when it follows its path; multiply swept areas count only once.

2.2 Measures for a Trajectory Amidst Other Trajectories

Some trajectory measures require the presence of other trajectories, for example measures for notions like similarity and centrality. For now we only consider other trajectories that exist at the same time as the trajectory that we observe. However, it is also possible to take other trajectories into account that occur later or earlier, by adapting the measures.

Similarity

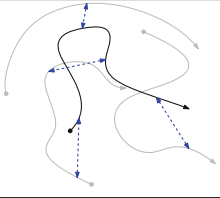
Description	<i>The average distance to another trajectory</i>		
Unit	meter	Range $ [0, \infty)$	
Derived from	-		
Related works	(Nanni and Pedreschi 2006)		

Similarity has been studied extensively. It is a distance measure between two trajectories that can refer to average distance between the entities, maximum distance between the entities, or maybe just similarity in shape of the path of the trajectory. Well-known similarity measures for paths of trajectories are the Hausdorff distance (Alt et al. 1995) and the Fréchet distance (Alt and Godau 1995). Well-known similarity measures for trajectories are the time-focused distance, the dynamic time-warping distance and the edit distance. All of these similarity measures take the location into account. However, one can easily define speed similarity or acceleration similarity if location is not relevant in the application. Another similarity mea-

sure, introduced for shapes, is the turn function similarity. It is rotation-invariant, and therefore particularly useful for comparing shape only.

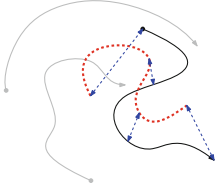
In this work, we define the similarity as the average distance between two trajectories during their time interval. To get the average distance, we can take the integral of the distance over the time interval and divide by the duration of the trajectories. Alternatively, one can take other possible options such as taking the minimum (or maximum) value, or integrating over space (reparameterize the trajectory).

Closeness

Description	<i>Average distance to the nearest other trajectory at each time</i>			
Unit	meter	Range	$[0, \infty)$	
Derived from	distance to the closest trajectory			
Related works	(Giardina 2008)			

The closest entity to a single moving entity at a single time can change over its time interval. Therefore, the average distance can show how close a trajectory is to other trajectories.

Centrality

Description	<i>Average distance to the central position of several trajectories</i>			
Unit	meter	Range	$[0, \infty)$	
Derived from	centrality			
Related works	(Agarwal et al. 2005; Laube et al. 2007)			

Centrality is also an intuitive concept that allows various different definitions. The centrality of an entity amidst other entities can change over time, so we first consider centrality measures at a fixed time. In other words, we have a set of points in the plane and one specific point p , and we want to describe how central p is. One option is to define the center of mass as the most central location, and use the distance of p to the center of mass to define p 's centrality. We can do something similar with the median location, whose coordinates are the median x -coordinate and the median y -coordinate of all points. Another option is to use the distance to the center of the smallest enclosing circle of the points. For one trajectory, we take the average of centrality over its time interval, which is defined by an integral. Similar with the two previous measures, other options than taking the average are also possible.

A more combinatorial notion of centrality is the minimum number of points of the set that must be removed to bring p to the convex hull of the points.

Discussion. Other notions that can be turned into measures are isolation and sociality. To define these, we could use nearest-neighbor distance or k -nearest neighbor distances.

2.3 Measures for a Trajectory in an Environment

Since the environment influences how an entity moves, it is natural to consider trajectories in the context of other data. Other data can be internal or external, that is, it can refer to the moving entity or to its environment. Depending on the source of the data, internal attribute data can be heart rate, brain activity, CO₂ level in exhaust, and produced noise. External attribute data can be current landcover, elevation, weather, quality of road, water currents, and visibility information.

Environmental factors can influence how the measures defined before may be redefined. For example, two trajectories at distance 80 m in the open field can be deemed more similar than two trajectories at 60 m, one of which is in the field and one in the forest. A speed of 10 m/s in the open field may be considered similar to a speed of 8 m/s in bushland.

Moreover, the environment may give rise to measures that do not exist without it, like a measure for the diversity of landcover types crossed by the moving entity. Here, the locations of the moving entity is used to select other data (the environment) to which the measure relates.

The definitions of measures for trajectories in environments are very much application-dependent, and therefore we omit further discussion in this paper that aims to take a general view on measures for trajectories and groups.

3 Measures for a Group of Trajectories

When sufficiently many moving entities are close enough for a relatively long period of time, they form a *group*. Closely related concepts are flocks, clusters, herds, and convoys. Formal definitions are given in (Gudmundsson and van Kreveld 2006; Jeung et al. 2008; Buchin et al. 2015; van Kreveld et al. 2016; Li et al. 2010; Kalnis et al. 2005). Here, we assume that a group is a fixed set of entities during a fixed period of time. Intuitively, the existence of a group relies on some conditions, such as a minimum number of entities in it, a minimum entity inter-distance, a minimum duration, etc. It is possible to add further requirements to a group definition, like similarity of heading of the group entities.

Just like for a single trajectory, many measures exist for a group of trajectories. Some of these are direct extensions from the single trajectory case, while others only occur for a group. In general, there are three views of treating a group when defining these measures:

Representative view: A single trajectory is used (not necessarily from the group) to define the measure. For example, we can take the mean location at every time and use this to define a mean trajectory. This mean trajectory can then be used to define group speed, for instance.

Complete view: All entities in the group are used to define the measure. For example, to define group speed in a different way than in the representative view, we can take the average speed of each entity over the group duration, and average this over the entities.

Area view: The area the group occupies is used to define the measure. For example, to define group density, we could define the area that the group occupies and let the density be the ratio of the number of entities and the area.

Not all views make sense for all measures we will define. Often, these different views give rise to different possible measures for the same concept, like for our example of group speed. The representative view gives us a way to generalize all single trajectory measures to groups.

In the following subsections we define group measures that do not require other input, group measures that exist for a group amidst other groups of moving entities, and group measures that require other types of input. The measures exist for any definition of a group, assuming that the number of entities in the group does not change in the interval during which this group exists. We also assume that the group exists during exactly one time interval, so we do not allow a group to form, break up, and later form again and call it the same group. Both assumptions can be lifted if required by the application at hand; they are made for the sake of clarity of the definitions of the measures.

3.1 Measures for a Single Group in Isolation

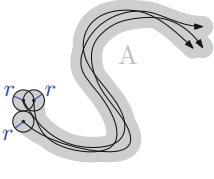
3.1.1 Measures Extended from a Single Trajectory

Almost all measures for a single trajectory in isolation can be extended directly to a group. Both the representative view and the complete view work well with these measures (except for the covered area). For example, the traversed distance of a group can be defined in two ways: as the average of the traversed distance of all entities in the group or as the traversed distance of a representative.

Furthermore, each of these two views can have more variations. With the complete view, instead of taking the average of the measure from each trajectory, we can also choose to take, for example, either the minimum or the maximum value. There are also several possibilities to get a representative trajectory. For example, to define the representative trajectory, we can take the mean or median point of all entities in a group at a single time, or just pick one trajectory to represent the group.

Only one measure for a single trajectory cannot be extended directly to a group:

Area covered

Description	<i>The union of the total area covered by each trajectory (using a disc of radius r) in a group (A)</i>		
Unit	square meter (m^2)	Range $[1, \infty)$	
Derived from	location		
Related works	(Boinski and Garber 2000)		

Clearly, we cannot sum up the covered area from all entities in a group because they might overlap. For the same reason, it is also not possible to aggregate the covered area by a group at each single time. Therefore, we define the area covered of a group as a union of the total area covered by its entities.

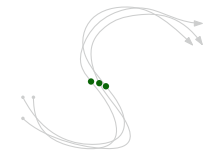
Note that this measure relies on a parameter r to define the disc which is used to sweep the covered area.

The movement behavior of a group influences this measure. For instance, when entities in a group move by following each other (*single file behavior*), then the group covers roughly the same area as one of its entities. A more compact group will also yield a smaller value for this measure than a more spread out group.


3.1.2 Measures that only Exist for a Group

In addition to attributes of a group that exist in a single trajectory, there are also attributes that only naturally exist when we observe multiple entities at a single time. For example, the number of entities in a group, (which is fixed for the whole duration of a group), or the *density* of a group.

Size

Description	<i>The number of entities in a group (n)</i>		
Unit	-	Range $[2, \infty)$	
Derived from	the number of entities in a group		
Related works	(Beauchamp 2012)		

Density

Description	<i>The average of the density of a group over its duration (ρ)</i>		
Unit	per square meter (m^{-2})	Range $ [\frac{1}{\pi r^2}, \infty)$	
Derived from	density		
Related works	(Peters and Krisp 2010; Beauchamp 2012)		

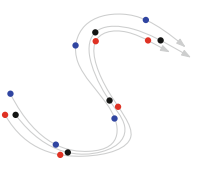
Intuitively, density relates a count to an area. Therefore, it is natural to define the density using an area view: a ratio of the size of a group and the covered area at a single time. Then, we integrate it over time to get a single value to represent the density of the group.

Other options to define the density of a group are based on the distance between entities in a group. We describe several possible definitions:

- The average of all distances between pairs of entities in a group.
- The weighted average of all distances, where small distances get a higher weight. Closer entities have more influence on the density than the farther ones.
- The ratio between the farthest distance occurring between two entities in the group and the maximum distance possible between two entities in any one group of the same size.

These distance-based measures have the advantage that they do not depend on a parameter to be determined.

Formation stability

Description	<i>The average of the magnitudes of the velocity difference of each entity and the group (F)</i>		
Unit	meter per second	Range $ [0, \infty)$	
Derived from	velocity		
Related works	(Heppner 1997; Viscido et al. 2004)		

A group of moving entities may move in a mostly stable formation, for instance, the V-formation of migrating birds. Often, entities keep the same formation for several reasons, like increasing the efficiency of movement by using the least energy.

At a fixed time, we define this measure using the magnitude of the difference in velocity of each entity and the group. We can average these magnitudes over the group entities and over time. Lower values for this measure indicate that the group has a low deformation rate. The formula to compute F for a group $G = \{e_1, e_2, \dots, e_n\}$ over the duration $[t_\alpha, t_\beta]$ is:

$$F = \frac{1}{t_\beta - t_\alpha} \int_{t_\alpha}^{t_\beta} \frac{1}{n} \sum_{i=1}^n \|e_i(t) - \mathbf{G}(t)\| dt$$

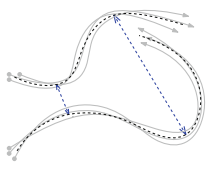
where $e_i(t)$ is the velocity vector of entity e_i at time t and $\mathbf{G}(t)$ is the velocity vector of G , which is the average over all of its entities at time t .

Discussion. Many more application-specific group measures are conceivable. For instance, the occurrence of a *leader*. A leader can be defined as a single entity that is usually in front of the rest of the entities in the group, or as an entity whose movement drives the movement of the group (Boinski and Garber 2000; Andersson et al. 2007). Some groups are more clearly led than others, which can be captured in a measure. Because such measures depend more on behaviour interpretation than a clear objective aspect of groups, we will not suggest any formal definition in the present article.

3.2 Measures for a Group Amidst Other Groups

Most of the measures for a group among other groups can be directly defined using the representative trajectory of each group and apply to measures for a trajectory amidst other trajectories.

Similarity

Description	<i>The average distance to another group</i>			
Unit	meter	Range	$[0, \infty)$	
Derived from	location			
Related works	(Bento 2016)			

To define similarity between two groups of entities we need to be aware of a few issues: (i) The two groups may have different size, and we still want a similarity measure that makes sense. (ii) The trajectories in one group may be dissimilar among each other, and if this occurs in the same way in the other group, then the groups are similar. These considerations suggest a many-to-many matching-based definition. For each trajectory in each group, consider the similarity to the most similar trajectory in the other group, and average over these similarity values. If the one group has n trajectories and the other m , then we use $n + m$ similarity values to average over. Alternatively, we could take all $n \cdot m$ pairs of trajectories, one from each group, and take their average, but this does not allow groups to be similar if their trajectories are not similar internally.

Alternatively, we can compare the similarity between the two groups at a single time and then integrate the result over the duration of the groups.

If we ignore the temporal component, then the area view can be used to compare the shape of the groups by comparing the shape of their covered area for a group shape similarity measure.

Closeness

Description	<i>The average distance to the nearest other group at each time</i>			
Unit	meter	Range	$[0, \infty)$	
Derived from	distance to the closest group			
Related works	-			

Although this measure is a straightforward extension from the measure for a single trajectory, applying it directly to the representative trajectories might not give the proper estimation about how close the two groups are. Intuitively, a group is near to the other groups if one or more entities in the group is close to entities in other groups. However, because the representative trajectory is located roughly in the middle of other entities in the group, there must be other pair of entities (each from different groups) that are closer.

Since the nearest other group might change over the duration of the group, it is better to first define the distance between two groups at a fixed time. Then, we can pick the minimum distance to one other group and get the average of this value over the whole duration of the group. At a single time, the distance between two sets of points can be defined in several ways. The simplest is to take the closest distance from any entity in one group to any entity in the other. Other options include the Hausdorff Distance (Alt et al. 1995) and the Earth Mover’s Distance (Rubner et al. 2000). The question is whether we consider two groups to be very close when they are completely interspersed or when they move closely side by side as a whole group.

Centrality

Description	<i>The average distance to the central position of other groups</i>			
Unit	meter	Range	$[0, \infty)$	
Derived from	centrality			
Related works	-			

Centrality can be defined using the representative trajectories and then apply the centrality measure for a single trajectory. We can also use the complete view of a group by computing the centrality of each entity w.r.t the central position of all entities and then average them over time to get the centrality of the group.

Discussion. So far in this section, we only considered measures that assume all groups start and end at the same time. It is also possible to define measures if we drop this assumption. For example, we can measure how many other groups exist (on average) during the lifetime of a group, and we can measure how similar a group moves with respect to any previously existing group.

3.3 Measures for a Group of Trajectories in an Environment

Similar to the case for a single trajectory, we can consider environmental factors when defining the measure of a group. For example, entities in a group may change their formation because they are moving in single file formation while crossing a river, and may change back to the previous formation on the opposite bank. In this case, we probably want to say that the formation is consistent if the formation is the same in each terrain type. We cannot use formation stability to capture this.

Different types of moving entities show different behavior when moving as a group, and this is influenced by different types of external factors. Consequently, integrating those factors into measures for a group depends on the type of moving entities that we want to analyze. Therefore, measures that take external factors into account are application-dependent, and general purpose measures are less easily defined.

4 Applications of Group Measures

With the vast increase of trajectory data, large collections of trajectories must be organized well and good tools for selection, visualization, and analysis must be available. We discuss how group measures may be used for these tasks. We assume that a group structure exists and has been computed.

Selection. To explore a collection of trajectories organized in groups, we may want to select the ten or twenty most important or characteristic groups. What is important about a group is application-dependent and can be discussed, but generally one can say that large groups with a long duration are more important than small groups with a short duration. Any of the other attributes can contribute to the importance of a group too: perhaps fast-moving groups (high average speed) or groups with a large area covered are important. The more measures play a role, the more dimensions the Pareto-optimal front has, and the more Pareto-optimal choices there are.

While the single most important group may be defined by weighing the different measures that contribute to importance, selecting a set of ten most characteristic and important groups is more complex. In many applications the selection should be *varied*, implying that choosing one group in the set influences which other groups should also be chosen. The corresponding issues have been studied in the classical

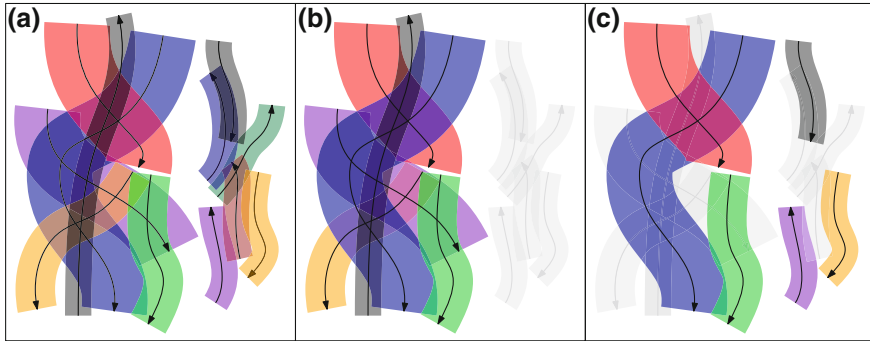


Fig. 1 **a** Each group is represented using a single *black* trajectory and an area with different colors. The width of the area shows the size of the group. **b** Only the six longest and biggest groups are selected, shown in color. **c** A better selection of six groups for diversity of the groups with respect to location.

problem of *settlement selection* (van Kreveld et al. 1997; Samsonov and Krivosheina 2012): not necessarily the ten biggest cities should be chosen, they should also be spatially distributed over the region of interest. Various models for such selections have been suggested in the literature. Transferring these issues to groups in collections of trajectories, it may be important to get variation in where in the plane (in space) the group occurred. If the six biggest and longest duration groups all occurred in the west, the next six ones occurred in the east, and we wish to select six groups, then it would probably be undesirable to select all groups in the west; instead, three or four in the west and three or two in the east gives a better idea of the data (see Fig. 1).

While a full treatment on when groups are important, salient or characteristic, and when a selection is an appropriate selection is beyond the scope of this paper, it is clear that group measures do play an important role in the process.

Visualization. Visualization of large collections of trajectories is difficult for several reasons. First of all, it is not easy to show time in an intuitive manner, besides animations. Second, simply showing all paths of trajectories usually creates a chaos in which little can be seen. If the trajectories come in groups, it is attractive to identify these groups and use visual encodings of relevant attributes or measures. A representative can be used to show each group (determined by trajectory measures), the group size can be visualized by line thickness of the representative, and the group speed or density can be visualized by color or other visual variable.

Segmentation. The segmentation of a single trajectory has been discussed in several papers (Mann et al. 2002; Buchin et al. 2011c; Aronov et al. 2013), but it may be more appropriate to *segment a group* in case the group as a whole shows certain behavior types. A herd of bison may be roaming, migrating, or sleeping. Segmentation is typically determined by within-segment similarity and across-segment dissimilarity and hence it relies on suitable measures. How to extend trajectory segmentation to

trajectory group segmentation is an interesting research question in which it is clear that group measures will play an important role.

5 Conclusions and Future Work

In this paper, we discussed measures for a single trajectory and a group of trajectories arising from moving entities. These measures provide extra information than just the spatial and temporal component needed to be able to analyze the trajectory data better. First, we gave basic measures for a single trajectory and differentiated them into three types: measures for a single trajectory, measures for a trajectory amidst other trajectories, and measures for a trajectory in the context of other data.

For a group of trajectories, we introduced three different views to define measures for a group: the representative view, the complete view and the area view. Most measures for a single trajectory can be extended directly to groups using at least one of the three views. We also introduced exclusive measures for a group like density and formation stability. For measures for group amidst other groups, we discussed their differences with the same measure for a single trajectory and gave alternative approaches to define them.

Finally, we considered several tasks where group measures are essential to analyze collections of trajectories data.

Our discussion of measures gives rise to various directions of future research. Firstly, we can consider real-world trajectory data and investigate the need for measures beyond the ones we have given. It is also interesting to analyze how external factors like geographic context should influence measures in various applications. Secondly, the three cases selection, visualization, and segmentation of groups have only been addressed briefly, and a complete study of each of these tasks would be valuable. Thirdly, we may examine the measures further on robustness to “outliers” within a group. Finally, measures need to be computed by algorithms, which in several cases still need to be developed.

Acknowledgements Lionov Wiratma is supported by the Ministry of Research, Technology and Higher Education of Indonesia (138.41/E4.4/2015). Marc van Kreveld is partially supported by the Netherlands Organisation for Scientific Research (NWO) under project “More Content with Geometric Content”. Maarten Löffler is partially supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 614.001.504.

References

- Agarwal PK, de Berg M, Gao J, Guibas LJ, Har-Peled S (2005) Staying in the middle: exact and approximate medians in R^1 and R^2 for moving points. In: Proceedings of the 17th Canadian conference on computational geometry, CCCG’05, University of Windsor, Ontario, Canada, 10–12 Aug 2005, pp 43–46

- Alt H, Godau M (1995) Computing the fréchet distance between two polygonal curves. *Int J Comput Geom Appl* 05(01n02):75–91
- Alt H, Behrends B, Blömer J (1995) Approximate matching of polygonal shapes. *Ann Math Artif Intell* 13(3):251–265
- Anagnostopoulos A, Vlachos M, Hadjieleftheriou M, Keogh EJ, Yu PS (2006) Global distance-based segmentation of trajectories. In: Proceedings of the twelfth ACM SIGKDD international conference on knowledge discovery and data mining, Philadelphia, PA, USA, 20–23 Aug 2006, pp 34–43
- Andersson M, Gudmundsson J, Laube P, Wolle T (2007) Reporting leadership patterns among trajectories. In: Proceedings of the 2007 ACM symposium on applied computing (SAC), Seoul, Korea, 11–15 Mar 2007, pp 3–7
- Andrienko GL, Andrienko NV, Bak P, Keim DA, Wrobel S (2013) *Visual analytics of movement*. Springer, Berlin
- Andrienko N, Andrienko G, Pelekis N, Spaccapietra S (2008) Basic concepts of movement data. In: Giannotti F, Pedreschi D (eds) *Mobility, data mining and privacy*, 1st edn. Springer
- Aronov B, Driemel A, van Kreveld MJ, Löffler M, Staals F (2013) Segmentation of trajectories for non-monotone criteria. In: Proceedings of the twenty-fourth annual ACM-SIAM symposium on discrete algorithms, SODA 2013, New Orleans, Louisiana, USA, 6–8 Jan 2013, pp 1897–1911
- Beauchamp G (2012) Flock size and density influence speed of escape waves in semipalmated sandpipers. *Anim Behav* 83(4):1125–1129
- Benhamou S (2004) How to reliably estimate the tortuosity of an animal's path: straightness, sinusosity, or fractal dimension? *J Theor Biol* 229(2):209–220
- Benkert M, Gudmundsson J, Hübner F, Wolle T (2008) Reporting flock patterns. *Comput Geom* 41(3):111–125
- Bento J (2016) A metric for sets of trajectories that is practical and mathematically consistent. *CoRR* abs/1601.03094
- Boinski S, Garber P (2000) *On the move: how and why animals travel in groups*. Nature/Science. University of Chicago Press
- Buard E, Brasebin M, IGN S (2011) Visual exploration of large animal trajectories. In: 25th international cartographic conference (ICC11). ICC 2011, 3–8 July 2011 Paris, France, p 7
- Buchin K, Buchin M, Gudmundsson J, Löffler M, Luo J (2011a) Detecting commuting patterns by clustering subtrajectories. *Int J Comput Geom Appl* 21(3):253–282
- Buchin K, Buchin M, van Kreveld MJ, Luo J (2011b) Finding long and similar parts of trajectories. *Comput Geom* 44(9):465–476
- Buchin M, Driemel A, van Kreveld MJ, Sacristán V (2011c) Segmenting trajectories: a framework and algorithms using spatiotemporal criteria. *J Spat Inf Sci* 3(1):33–63
- Buchin K, Buchin M, van Kreveld MJ, Speckmann B, Staals F (2015) Trajectory grouping structure. *JoCG* 6(1):75–98
- Calenge C, Dray S, Royer-Carenzi M (2009) The concept of animals' trajectories from a data analysis perspective. *Ecol Inform* 4(1):34–41
- Dodge S, Weibel R, Lautenschütz AK (2008) Towards a taxonomy of movement patterns. *Inf Vis* 7(3):240–252
- Giardina I (2008) Collective behavior in animal groups: theoretical models and empirical studies. *HFSP J* 2(4):205–219
- Giuggioli L, Potts JR, Harris S (2011) Animal interactions and the emergence of territoriality. *PLoS Comput Biol* 7(3):1–9
- Gudmundsson J, van Kreveld MJ (2006) Computing longest duration flocks in trajectory data. In: Proceedings of 14th ACM international symposium on geographic information systems, ACM-GIS 2006, 10–11 Nov 2006, Arlington, Virginia, USA, pp 35–42
- Gudmundsson J, van Kreveld MJ, Staals F (2013) Algorithms for hotspot computation on trajectory data. In: 21st SIGSPATIAL international conference on advances in geographic information systems, SIGSPATIAL 2013, Orlando, FL, USA, 5–8 Nov 2013, pp 134–143

- Hanks EM, Hooten MB, Johnson DS, Sterling JT (2011) Velocity-based movement modeling for individual and population level inference. *PLoS ONE* 6(8):1–17
- Heppner F (1997) Three-dimensional structure and dynamics of bird flocks. In: Parrish JK, Hamner WM (eds) *Animal groups in three dimensions: how Species Aggregate*, 1st edn. Cambridge University Press
- Huang Y, Chen C, Dong P (2008) Modeling herds and their evolvments from trajectory data. In: Cova TJ, Miller HJ, Beard K, Frank AU, Goodchild MF (eds) *Geographic information science, Proceedings of the 5th international conference, GIScience 2008, Park City, UT, USA, 23–26 Sept 2008. Lecture notes in computer science*, vol 5266. Springer, pp 90–105
- Hwang S, Liu Y, Chiu J, Lim E (2005) Mining mobile group patterns: A trajectory-based approach. In: Ho TB, Cheung DW, Liu H (eds) *Advances in knowledge discovery and data mining, Proceedings of the 9th Pacific-Asia conference, PAKDD 2005, Hanoi, Vietnam, 18–20 May 2005. Lecture notes in computer science*, vol 3518. Springer, pp 713–718
- Jeung H, Yiu ML, Zhou X, Jensen CS, Shen HT (2008) Discovery of convoys in trajectory databases. *PVLDB* 1(1):1068–1080
- Kalnis P, Mamoulis N, Bakiras S (2005) On discovering moving clusters in spatio-temporal data. In: Medeiros CB, Egenhofer MJ, Bertino E (eds) *Advances in spatial and temporal databases, Proceedings of the 9th international symposium, SSTD 2005, Angra dos Reis, Brazil, 22–24 Aug 2005. Lecture notes in computer science*, vol 3633. Springer, pp 364–381
- van Kreveld M, van Oostrum R, Snoeyink J (1997) Efficient settlement selection for interactive display. In: *Auto-Carto XIII Proceedings of the International Symposium on Computer-Assisted Cartography*, pp 287–296
- van Kreveld M, Löffler M, Staals F, Wiratma L (2016) A refined definition for groups of moving entities and its computation. In: *Proceedings of the 27th international symposium on algorithms and computation, ISAAC 2016, 12–14 Dec 2016, Sydney, Australia*, pp 48:1–48:12
- LaPoint S, Gallery P, Wikelski M, Kays R (2013) Animal behavior, cost-based corridor models, and real corridors. *Landsc Ecol* 28(8):1615–1630
- Laube P, Dennis T, Forer P, Walker M (2007) Movement beyond the snapshot dynamic analysis of geospatial lifelines. *Comput Environ Urban Syst* 31(5):481–501
- Lee J, Han J, Whang K (2007) Trajectory clustering: a partition-and-group framework. In: *Proceedings of the ACM SIGMOD international conference on management of data, Beijing, China, 12–14 June 2007*, pp 593–604
- Lee J, Han J, Li X (2008a) Trajectory outlier detection: a partition-and-detect framework. In: *Proceedings of the 24th international conference on data engineering, ICDE 2008, 7–12 Apr 2008, Cancún, México*, pp 140–149
- Lee J, Han J, Li X, Gonzalez H (2008b) TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *PVLDB* 1(1):1081–1094
- Li Z, Ding B, Han J, Kays R (2010) Swarm: mining relaxed temporal moving object clusters. *PVLDB* 3(1):723–734
- Liu H, Schneider M (2012) Similarity measurement of moving object trajectories. In: *Proceedings of the third ACM SIGSPATIAL international workshop on GeoStreaming, ACM, New York, NY, USA, IWGS '12*, pp 19–22
- Mann R, Jepson AD, El-Maraghi TF (2002) Trajectory segmentation using dynamic programming. In: *16th international conference on pattern recognition, ICPR 2002, Quebec, Canada, 11–15 Aug 2002*, pp 331–334
- Nanni M, Pedreschi D (2006) Time-focused clustering of trajectories of moving objects. *J Intell Inf Syst* 27(3):267–289
- Peters S, Krisp JM (2010) Density calculation for moving points. In: *Proceeding of the 13th AGILE international conference on geographic information science, Guimaraes, Portugal, 11–14 May 2010*, pp 43–46
- Ranacher P, Tzavella K (2014) How to compare movement? a review of physical movement similarity measures in geographic information science and beyond. *Cartogr Geogr Inf Sci* 41(3)

- Rubner Y, Tomasi C, Guibas LJ (2000) The earth mover's distance as a metric for image retrieval. *Int J Comput Vis* 40(2):99–121
- Safi K, Kranstauber B, Weinzierl R, Griffin L, Rees EC, Cabot D, Cruz S, Proaño C, Takekawa JY, Newman SH, Waldenström J, Bengtsson D, Kays R, Wikelski M, Bohrer G (2013) Flying with the wind: scale dependency of speed and direction measurements in modelling wind support in avian flight. *Mov Ecol* 1(1):4
- Samsonov T, Krivosheina A (2012) Joint generalization of city points and road network for small-scale mapping. In: Proceedings of seventh international conference on geographic information science GIScience, Columbus, Ohio, USA, 18–21 Sept 2012, pp 18–21
- Viscido SV, Parrish JK, Grünbaum D (2004) Individual behavior and emergent properties of fish schools: a comparison of observation and theory. *Mar Ecol Prog Ser* 273:239–249
- Yuan J, Zheng Y, Zhang C, Xie W, Xie X, Sun G, Huang Y (2010) T-drive: driving directions based on taxi trajectories. In: Proceedings of the 18th ACM SIGSPATIAL international symposium on advances in geographic information systems, ACM-GIS 2010, 3–5 Nov 2010, San Jose, CA, USA, pp 99–108