JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 24 (2022), pp. 1–30 doi:10.5311/JOSIS.2022.24.160

RESEARCH ARTICLE

An experimental evaluation of grouping definitions for moving entities*

Lionov Wiratma[1], Marc van Kreveld[2], Maarten Löffler[2], and Frank Staals[2]

[1]Dept. of Informatics, Parahyangan Catholic University, Indonesia
[2]Dept. of Information and Computing Sciences, Utrecht University, the Netherlands

Received: December 2, 2020; returned: February 24, 2021; revised: September 3, 2021; accepted: February 1, 2022.

Abstract: One important pattern analysis task for trajectory data is to find a *group*: a set of entities that travel together over a period of time. In this paper, we compare four definitions of groups by conducting extensive experiments using various data sets. The grouping definitions are different by one or more of three different characteristics: whether they use the measured sample points or continuous movement, how distance is used to decide if entities are in the same group, and whether the duration of the group is measured cumulatively or as one contiguous time interval. We are interested in the differences between the definitions and comparisons to human-annotated data, if available. We concentrate on pedestrian data and on different crowd densities. Furthermore, we analyze the robustness of the definitions with respect to their dependence on different sampling rates. We use two types of trajectory data sets: synthetic trajectories and real-life trajectories extracted from video surveillance. We present the results of the quantitative evaluations. For experiments with real-life trajectories, we augment them with a qualitative evaluation using videos that show groups in the trajectories with a color coding.

Keywords: trajectories, collective motion, groups, experimental comparison

*Preliminary versions of this paper appeared in Proc. of the 10th International Conference on Geographic Information Science (GIScience 2018) and Proc. of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL 2019.

1 Introduction

Movement data of a single moving entity is typically described as a *trajectory*. Formally, a trajectory is a continuous mapping from a time interval $\mathcal{I} = [t_{start}, t_{end}]$ to the space in which the entity is moving. Even though the movement is often continuous, a tracking device usually reports the location of an entity location only at specific moments with regular or irregular intervals in between. Therefore, trajectory data is often stored as an ordered sequence of discrete time-stamped locations. For example, trajectory $T = \{(p_1, t_1), (p_2, t_2), \ldots, (p_\tau, t_\tau)\}$ represents the movement of an entity, where $p_i = (x_i, y_i)$ denotes the position of the entity in a two-dimensional space at time t_i and τ is the total number of stored data points. Since the original movement is continuous, we must assume a position at any time between any two data points, and linear interpolation (constant velocity) is the simplest assumption.

There are many different ways to analyze movement data. Especially interesting are the patterns that involve interactions between the entities, like leadership [1], and chasing/avoidance behavior [8, 25]. Collective movement patterns in which multiple entities travel together during a period of time have been extensively studied within this class. There are multiple applications. In ecology, researchers try to understand the behavior of groups of animals [18, 32]. In veterinary science, researchers investigate whether the composition of animals in a group depends on the health level of its members [7, 9]. In social psychology researchers analyze crowds to analyze human behavior [35]. In security systems, suspect behavior may be identified by the way that individuals move together. In all these areas, identifying collective movement in trajectory data can provide critical new insights.

Many different definitions have been suggested to model the collective movement of a "sufficiently large" set of entities that travel "together" for a "sufficiently long" period of time: flocks [4,13,44], mobile groups [16], moving clusters [21], moving micro-clusters [27], herds [15], convoys [19], swarms [28], gatherings [55], traveling companions [42], platoons [26], groups [5], refined groups [43], crews [29], and evolving companions [37].

It is beyond the scope of this paper to review them all and explain their often subtle differences. Refer to the original papers for details. Most of these papers (i) introduce a new collective movement, (ii) present one or more algorithms to compute it, and (iii) describe experiments where the new algorithms are run on some data sets (sometimes comparing the results to one earlier type). However, no extensive experimental study that includes several of such definitions and analyzes their differences has been performed. We believe that such a comparative study is more useful than yet another definition of groups.

Our Contribution In this paper, we provide an extensive experimental study to find small groups in pedestrian data. This is an important case in analyzing the throughput in public spaces like shopping malls [38], parks [23, 50] and train stations [36], and in detecting suspect behavior in such spaces [10, 20]. With the Covid-19 pandemic, the application to identifying possible transmission of a disease has become highly relevant as well. Small groups can be as small as just two individuals.

We compare four of the definitions, namely *convoys* [19] (which in our setting are the same as traveling companions [41]), *swarms* [28], *original groups* [5], and *refined groups* [43]. These four definitions differ in

input: how they model the input (as a continuous function or by discrete time stamps),

connectivity: how they model when entities are considered together, and
duration: how they measure if the entities are together long enough.

Convoys (and also traveling companions) are a well-known type that considers groups whose composition does not change, and assessment is done at the time stamps themselves. Original groups and refined groups distinguish themselves from the other definitions on the input aspect: they treat time as a continuous phenomenon. This may be more accurate than the discrete sampling based definitions when the sampling of the trajectory is relatively low. The refined groups definition is the only definition that measures togetherness within the group only, not having other non-group entities influence this. Swarms distinguish themselves by not requiring a contiguous grouping; interruptions are allowed. We discuss these four definitions in more detail in the next section.

Early definitions that use a shape of the cluster (flocks) in the definition are disregarded because they exhibit the lossy flock problem [19]. Since we study small groups, allowing group composition change is not suitable and hence we do not take definitions into account where the composition may change (like gatherings). We also exclude definitions that focus more on how moving entities converge before becoming a group [54]. Some other definitions are motivated mostly by vehicle data (e.g. evolving companions) we also do not consider. Finally, we note that the four chosen definitions all use three main parameters: one for group size, one for group duration, and one for inter-distance. Therefore, comparing these definitions is more clean than including more complex definitions that need more parameters (e.g., herds, moving micro-clusters, platoon, crews, evolving companions). There are no definitions that use fewer parameters. Note that convoys and swarms use one additional parameter for the density threshold, but we fix its value. We provide a full explanation of this parameter in the next section.

Our study is not just a comparison of the four definitions (i.e., convoys, swarms, original groups, and refined groups), we also investigate of how the input, space, and time can be treated and how this affects the results of the experiments. The objective of our study is not to find the "best" definition since we typically do not have the ground truth. However, some data sets provide human annotations of groups, and we will compare the four definitions to this. Unfortunately, no information is provided on how groups are annotated, and we realize that the annotation is subjective by its nature. So even when we compare the four definitions with human annotation, we can not identify a best definition, only a better correspondence with human intuition on how groups look.

We study the following research questions:

1. How well do the above definitions correspond with what humans consider a "group", and how do the characteristics mentioned (input, connectivity, and duration) influence this?
2. How does the number of groups, as reported by the various definitions, depend on the density of the entities?
3. How does the number of groups, as reported by the various definitions, depend on the sampling rate of the input trajectories?

We answer these questions by performing both a quantitative and qualitative analysis. For the quantitative analysis, we use real-life trajectory data sets to compare the reported groups by the four definitions with the human annotation groups. To evaluate the results, we use three evaluation metrics: *precision*, *recall*, and *F1-scores*.

We analyze the four definitions qualitatively only for the experiments with real-life trajectory data sets. The qualitative evaluation is aided by augmenting video footage. We introduce a visualization that shows the reported groups with color-coding in the video footage. Furthermore, the movement traces of the pedestrians in the video are also shown. This visualization allows easy comparison between the reported groups according to one definition and according to human-annotated groups, or according to groups from another definition.

Results and Organization. In the following section, we review the four grouping definitions that we consider and analyze how they differ in theory. Then, we describe the methods for our experimental comparison and introduce our new visualization method in Section 3. We present the results of our experimental evaluation in Section 4, before concluding the paper in Section 5.

2 The Definitions

The four definitions, Original Groups (OG), Refined Groups (RG), Convoys (CO), and Swarms (SW), rely on three parameters to define a group: the size parameter (the number of entities in a group), the temporal parameter (the time interval in which those entities form a group), and the spatial parameter (the distance between entities in the group). More formally, let $\mathcal{X}$ be the set of moving entities. A subset of entities from $\mathcal{X}$ forms a group $\mathcal{G}$ during time interval $\mathcal{I}$ when

- $\mathcal{G}$ contains at least m entities,
- $\mathcal{I}$ has a duration at least δ, and
- every pair of entities $x, y \in \mathcal{G}$ is *connected* during $\mathcal{I}$.

The size parameter, the required minimum of entities to form a group, has the same interpretation in all four definitions.

For the temporal parameter, the swarms [28] definition is it different from the others since it measures the duration of a group cumulatively. Let $\mathcal{T}$ be a set of timestamps where at each timestamp, every pair of entities in $\mathcal{G}$ are connected. Then, swarm uses the size of $\mathcal{T}$—the number of timestamps—to define the duration of $\mathcal{G}$, rather than the duration of one contiguous time interval $\mathcal{I}$. Note that with this property, swarm allows entities in $\mathcal{G}$ to leave $\mathcal{G}$ and join again later, as long as $\mathcal{G}$ is formed during at least δ timestamps (which may be non-consecutive). For the other three definitions, the duration of being together must be uninterrupted, but there is still a difference in the treatment of time. The original and refined group definitions consider trajectories in their continuous form and interpolate positions between timestamps using the linear interpolation. Therefore, the start and end time of the duration of a group will typically not be at any timestamp. For convoys, being together is considered only at the given timestamps themselves; all trajectories are assumed to have the same timestamps.

For the spatial parameter, we take a closer look at each definition. The original groups definition uses ε-connectivity between two entities as follows [5]: two entities x and y ($x, y \in \mathcal{X}$) are *directly ε-connected* if at any particular time t (not necessarily at a timestamp: the time when the position of an entity is recorded), the Euclidean distance between x and y is at most ε (here, $\varepsilon > 0$ is the spatial parameter). Furthermore, x and y are *ε-connected* *in* $\mathcal{X}$ at time t if there is a sequence $x = x_0, ..., x_k = y$, with $x_0, ..., x_k \in \mathcal{X}$ and for all i, x_i and x_{i+1} are directly ε-connected at time t. Note that to decide if, at time t, two entities are ε-connected we need only the locations of the entities at time t.

One may claim that it is more natural if the connectivity for x and y at time t can only be provided by *entities who are in the same group*, which is the approach taken by the refined group definition [43]. More specifically, to decide if x and y are ε-connected in a group $\mathcal{G}$, we ignore all entities not in $\mathcal{G}$ and require a sequence $x = x_0, ..., x_k = y$, with $x_0, ..., x_k \in \mathcal{G}$ where x_i and x_{i+1} are directly ε-connected at time t. Computing groups using this refined definition appears more complex than in the original groups case since we cannot decide just from the locations at time t whether x and y are ε-connected. We need the location history and future of these entities as well.

	Input	Connectivity	Duration
Original Groups (OG)	continuous	free	consecutive
Refined Groups (RG)	continuous	within group	consecutive
Convoys (CO)	discrete	free	consecutive
Swarms (SW)	discrete	free	cumulative

Table 1: Differences between the four grouping definitions.

The convoy [19,42] and swarm definitions [28] treat connectivity similar to the original groups, in that to determine connectivity at timestamp t, they need only the locations at time t. Note that these definitions consider the input as discrete, and hence connectivity is only defined at timestamps. Their notion of connectivity is slightly more general than the one considered by the original groups. They use the concept of density connection [11]. Let the ε-neighborhood $\mathcal{N}_\varepsilon(x)$ of an entity $x \in \mathcal{X}$ be the number of other entities in $\mathcal{X}$ that have the Euclidean distance at most ε ($\varepsilon > 0$) from x (at some given timestamp t). Now, given a density threshold μ ($\mu > 0$), an entity $y \in \mathcal{X}$ is *directly density-reachable* from x if $y \in \mathcal{N}_\varepsilon(x)$ and $|\mathcal{N}_\varepsilon(x)| \geq \mu$. Furthermore, y is *density-reachable* from x if a sequence of entities $\in \mathcal{X}$ exists where each consecutive pair of entities in the sequence from x to y is directly density-reachable. Clearly, if $\mu = 1$ then the notion of (directly) density-reachable is exactly the same as the (directly) ε-connected in the original group definition. Henceforth, we only use $\mu = 1$ since $\mu > 1$ prevents the convoy and swarm definitions identifying groups that contain only two entities (recall that our experiments focus on finding small groups). Furthermore, this choice makes the four grouping definitions better comparable, because they use the same three parameters, and our study can focus on the differences in the way these three parameters are treated in the definitions.

We summarize the differences between the four definitions in Table 1. We note that no two of the four definitions we consider are the same on all three aspects. Furthermore, we do not consider the four other combinations of the three aspects, since such definitions either do not exist or are more complex than using three parameters.

There are no definitions that consider continuous trajectories as input and count the duration cumulatively. To process trajectory data in its continuous form, we apply an interpolation between timestamps, and therefore, treat time consecutively. Of course, any group with the same entities can be formed at different times, and we can merge those durations cumulatively. However, this is a mix of consecutive and cumulative duration.

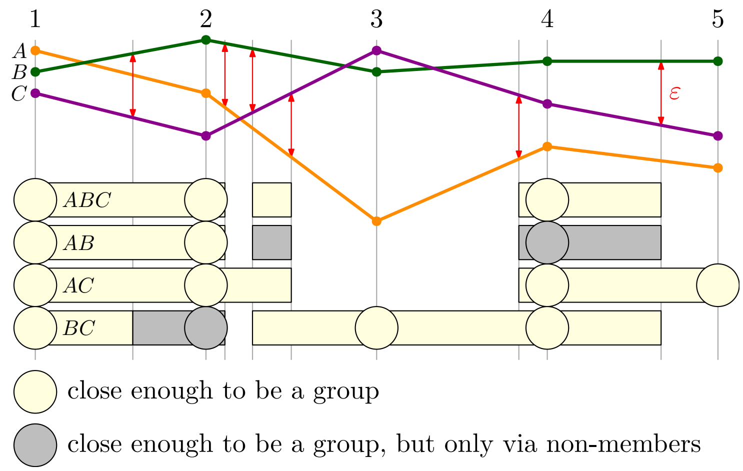

Figure 1: Maximal groups according to: ($m = 2, \delta = 1$)
Convoys(CO): $ABC[1, 2], ABC[4], AC[4, 5], BC[1, 4]$
Swarms(SW): $ABC(1, 2, 4), AC(1, 2, 4, 5), BC(1, 2, 3, 4)$
Original Groups(OG)/Refined Groups(RG): $ABC[1, 2.1], AC[1, 2.5], AC[3.8, 5], BC[2.3, 4.6]$

Some definitions consider discrete trajectories and use "within-group" connectivity. One example is platoon [26]. However, platoon uses a combination of consecutive and cumulative time duration. Other definitions define "togetherness" with slight variations. For example, the loose traveling companions [33] do not need to use the "within-group" connectivity for the entire duration of a group. Besides distance between entities, crews [29] also use many more aspects (e.g. speed, tortuosity, etc) to define the connectivity between entities. Consequently, these differences mean that gouping definitions need additional parameters. Therefore, we exclude them in the experiments.

Maximal Groups. In the original and refined group definitions [5, 43], a group $\mathcal{G}$ is a *maximal group* during time interval $\mathcal{I}$ if there is no time interval $\mathcal{I}' \supset \mathcal{I}$ for which $\mathcal{G}$ is also a group and there is no $\mathcal{G}' \supset \mathcal{G}$ that is also a group during $\mathcal{I}$. The swarms definition is extended in exactly the same way in [28], where it is called a *closed swarm*. The definition of convoys includes maximality by default, that is, only maximal convoys are convoys. Henceforth, we also use the term maximal group to describe the (maximal) convoy and the closed swarm.

Figure 1 illustrates the concept for a small example. Note that the same set of entities can appear multiple times (at different moments in time) as a maximal group under all definitions except swarms; for swarms it would be considered a single swarm with longer (summed) duration. Furthermore, all definitions allow an entity to be part of different groups, convoys, or swarms at the same time.

Differences. The differences shown in Table 1 affect how each definition specifies maximal groups from a set of trajectories. We demonstrate this using examples. First, we present an example in Figure 2, where a maximal group containing exactly the same entities may

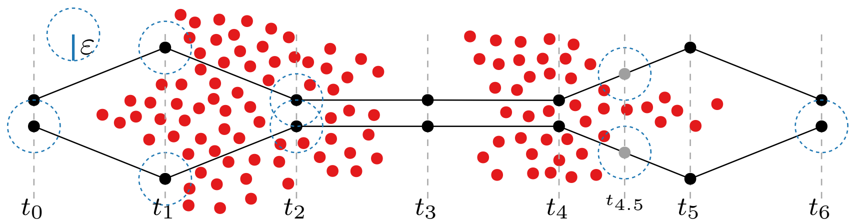

Figure 2: According to different definitions, the black entities are a group at different times.

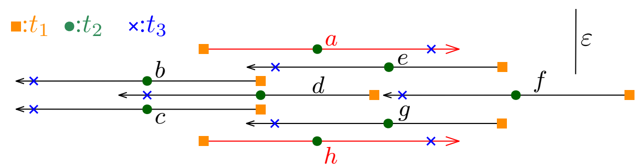

Figure 3: Entities a and h are not a refined group during $[t_1, t_3]$, but they are an original group, convoy, and swarm during $[t_1, t_3]$ or $\{t_1, t_2, t_3\}$ [43].

have different time durations, depending on which definition we use. Let two black entities x and y be the only entities that move; all red entities are stationary. Furthermore, trajectories of x and y consist of the shown positions at $t_0, t_1, ..., t_6$, and we set $\delta = 2$. According to the various definitions, the set $\{x, y\}$ is a group during

$\{t_0, t_1, t_2, t_3, t_4, t_6\}$ for swarms, since the only timestamp at which x and y are not ε-connected is t_5,

$[t_0, t_4]$ for convoys, since timestamps have to be consecutive (and the interval $[t_6, t_6]$ itself is too short),

$[t_1, t_{4.5}]$ for original groups, since the entities actually already stop being ε-connected at some intermediate time $t_{4.5}$ between timestamps t_4 and t_5, and

$[t_2, t_4]$ for refined groups, since x and y have to be connected using entities only from $\{x, y\}$ itself, they cannot connect using the red entities, and thus they form a refined group only when the distance between x and y is at most ε.

Next, we show that the type of connectivity between entities in a group, such as in the refined group definition can result in a completely different grouping. In particular, we use the same example provided by van Kreveld et al. [43]. In Figure 3, two entities a and h are moving in the same direction, opposite to the other entities. At any time during the time interval $\mathcal{I} = [t_1, t_3]$, a and h are ε-connected through other entities. As a consequence, the convoy and swarm definitions consider $\{a, h\}$ to be a group at timestamps t_1, t_2, t_3, or during interval $\mathcal{I}$ for the original group definition. In the refined group definition, $\{a, h\}$ is not a group during $\mathcal{I}$ because their connectivity is only through (changing) entities not in the group itself. Note that there are several refined groups that include a and h that have a considerably shorter duration.

3 Methods

To answer the research questions described previously, we conduct extensive experiments by computing all maximal groups from various trajectory data sets according to the different group definitions. We evaluate the results both quantitatively and qualitatively.

Data sets. To conduct our experiments, we use various data sets which we divide into three *categories* based on their source:

- Real-life trajectories extracted from video surveillance in a public area: the *NYC Grand Central Terminal* [51,52], *ETH Walking Pedestrian* [12,34], *Crowds by Examples* [24,39,40] and *Vittorio Emanuele II Gallery* data sets [3,39,40].
- Real-life trajectories of pedestrians walking in a laboratory environment: the *Pedestrian Dynamics* data set [17,53].
- Artificial trajectories generated by a computer simulation: the *Netlogo Flocking* data set [45,46].

We describe each data set in more detail along with the results of experiments using them in Section 4. The real-life data sets are captured from video surveillance; hence their raw coordinates are frame (pixel) coordinates from the videos. These coordinates are first converted to world coordinates using a homography matrix to be able to make fair distance comparisons. Most real-life trajectory data sets also come with a list of *human-annotated* groups; only the *NYC Grand Central Terminal* and the *Pedestrian Dynamics* data set do not.

Implementations. To compute all maximal groups according to the different notions of groups, we implement all algorithms ourselves:

- the Smart-and-Closed algorithm [42] to compute *convoys (traveling companions)*,
- the ObjectGrowth algorithm [28] to compute *swarms*,
- our implementation from Buchin et al. [5] to compute *original groups*, and
- our implementation from Wiratma et al. [48] to compute *refined groups*.

Note that since the swarm algorithm has an exponential running time, we were unable to compute all swarms for some of the parameter values in our experiments.

Quantitative Evaluation. We analyze and evaluate the results from all experiments quantitatively. We compute and count all maximal groups in our data sets according to the four definitions while varying the parameters of the definitions: the distance ε, the minimum time duration δ, and the minimum group size m.

- For several data sets, we provide the precision, recall, and F1-score as measurements to show the relevance between the groups found by each definition with the human-annotated groups. Note that we avoid the terms "correctness" and "ground truth": we can test only to what extent the groups found agree with the human-annotated groups. In particular, human-annotated data is likely to be influenced by personal interpretation and therefore not a ground truth.
- We vary the *density* of the environment by considering different numbers of entities moving in the same bounded space. We compute the number of groups for each definition and study how it changes with the number of entities.
- We vary the sampling rate, or level of detail, of the trajectories by ignoring a fraction of the vertices in each trajectory. We count how many groups are identified by the different definitions, and analyze the consistency of these numbers.

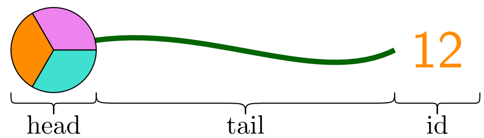

Figure 4: A moving entity is shown in a schematic manner.

Qualitative Evaluation. We also qualitatively evaluate the results of our experiments by visualizing the trajectories of pedestrians integrated in the videos from the data sets.

Conceptually, we represent each moving entity by a glyph that is overlaid on the video material; refer to Figure 4. Each glyph consists of three parts. The *head* is a disk which shows the *current location* of the represented entity. The *tail* is a piece of curve which shows the *previous locations* of the entity during a set duration. The *id* is a unique identifier of the entity.

Grouping information is encoded by the *color* of the head and tail. We use the color of the *tail* to show a human-annotated grouping. Entities belonging to the same group have the same color, and every entity can only belong to at most one group, which cannot change over time. Entities that do not belong to any group in the human-annotated grouping have a white tail. The color of the *head* indicates the grouping as computed by the method currently under study. The computed groups can in principle overlap, and they do change over time. As a result, the head of an entity can have multiple colors, and the color of a head may change as time progresses.

The combination of colors of the tails and heads gives insight into the matching between annotated groups and groups based on a grouping definition. Note that colors are chosen at random; even when a method produces a group that exactly matches with one of the annotated groups, the color of the head may be different than the color of the tail.

We applied this scheme to all our data sets and generated videos for various parameter settings; see Figure 5 for an example. Our implementation of the visualization is based on the work by Maurice Marx [30]. In the remainder of this paper, we supply some snapshots of interesting configurations. The complete collection of videos from this paper can be found on our website [47].

4 Experimental Evaluation

In this section we evaluate the results of our experiments. We focus our evaluation on the differences of the four definitions, and thus on the maximal groups that are reported, rather than the differences between the algorithms and their implementation. All implementations are non-optimized prototypes and therefore, comparing statistics like running time is meaningless.

4.1 Comparisons with Annotated Groups

We aim to establish how well the definitions capture the human intuition of grouping. To this end, we compute the groups, as reported by the various definitions, and compare them to human annotations. We then report the *precision* (the percentage of the groups according to the definition that also occur in the annotated data), the *recall* (the percentage

Figure 5: A snapshot from a video shows the grouping information. Observe the person (ID: 376) with two colors head (light blue and purple) in the middle. According to the human-annotated data, he only formed a group with another person next to him (ID 374), thus, they have the same tail color. However, a grouping definition considers the person as a member of two groups: a group of four persons with light blue heads and a group of five persons with purple heads.

of the human-annotated groups which are also groups according to the definition), and the corresponding *F1*-score. The F1-score is the harmonic mean of the recall (R) and the precision (P), and can be computed using the following formula: $F1 = 2\frac{P \cdot R}{P+R}$.

We use four data sets consisting of real-life trajectories from video surveillance: ETH Walking Pedestrian (ETH and HTL) [12], Vittorio Emanuele II Gallery (VEIIG) [39], and Crowds by Example (CBE) [39]. See Table 2 for details of each data set. Besides trajectories of pedestrians, these data sets are supplemented with homography matrices and lists of groups that are annotated manually by the authors. The annotations specify only *which* entities appear in a group, not *when*, or *how long* the entities form a group. Moreover, unlike in the four definitions, an entity occurs in at most one group in the human annotation.

For each definition, we count how many maximal groups match exactly with the annotated groups and evaluate the correctness using the precision, recall, and F1-score. For each data set we set the minimum required number of entities m to 2. The values for the inter-entity distance ε are chosen based on a study by Solera et al. [40], who analyze the average distance between people in the same group in a human crowd. Finally, we determine three different values for required minimum time δ that a group is together, based on the group annotations. In particular, we assume that a set of people cannot form a group when not all members are present in the video. Hence, we compute the time interval during which all members of an annotated group are present, and define the duration of the group to be the length of this interval. The minimum such duration over all groups gives us one choice of δ. The other two are chosen based on the average such duration $\bar{\delta}$ and the standard deviation σ. In particular, we pick $\delta = \bar{\delta} - \sigma$ and $\delta = \bar{\delta} - \frac{1}{2}\sigma$.

data set		ETH	HTL	VEIIG	CBE
input	video length	08:39	12:54	05:00	03:36
statistics	FPS	25	25	8	25
	#entities	360	389	630	434
	avg τ	143.37	159.2	189.82	400.12
annotation	g	58	39	207	115
statistics	G	2-6	2-3	2-7	2-4
experiment	ε	1.24	0.94	0.963	{1.22,1.52}
parameters	δ	{72,89,105}	{20,58,96}	{17,38,58}	{36,57,78}

Table 2: Information on the trajectories in the data sets and parameters used in the experiments. Here, g denotes the number of annotated groups, and G their size range. The video length is specified in minutes and seconds; the values for ε and δ are in meters and frames, respectively.

4.1.1 The ETH Walking Pedestrian.

The data set contains two bird-eye view videos from two locations: in front of a university building at ETH Zurich (ETH) and a sidewalk near a tram stop (HTL). We consider this data sets to have low to medium density. The results of the experiments using the ETH and HTL data sets are presented in Table 3 and Table 4, respectively.

		Precision	Recall	F1-score
$\delta = 72$	OG	0.410	0.741	0.528
	RG	0.467	0.741	0.573
	CO	0.453	0.741	0.562
	SW	0.380	0.793	0.514
$\delta = 89$	OG	0.411	0.638	0.500
	RG	0.463	0.638	0.537
	CO	0.446	0.638	0.525
	SW	0.372	0.724	0.491
$\delta = 105$	OG	0.418	0.569	0.482
	RG	0.471	0.569	0.515
	CO	0.452	0.569	0.504
	SW	0.404	0.655	0.500

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 3: Comparative results on the ETH data set.

As expected, if the density of the crowd is relatively low, all definitions have similar results, which we can see from their F1-score from both data sets. Furthermore, as we increase the value of δ, the recall values are decreased because now many annotated groups with a short duration cannot be found by all definitions.

We observe that swarm can match more annotated groups than other definitions but with a low precision because of swarm output more maximal groups. The other definitions output fewer maximal groups (and different between each definition), but the number of annotated groups that they are able to match is the same. Here, the strict connectivity requirement makes the refined definition gets better precision scores.

		Precision	Recall	F1-score
$\delta = 20$	OG	0.400	0.974	0.567
	RG	0.418	0.974	0.585
	CO	0.409	0.974	0.576
	SW	0.463	0.974	0.628
$\delta = 58$	OG	0.660	0.897	0.760
	RG	0.673	0.897	0.769
	CO	0.660	0.897	0.760
	SW	0.706	0.923	0.800
$\delta = 96$	OG	0.690	0.744	0.716
	RG	0.707	0.744	0.725
	CO	0.707	0.744	0.725
	SW	0.705	0.795	0.747

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 4: Comparative results on the HTL data set.

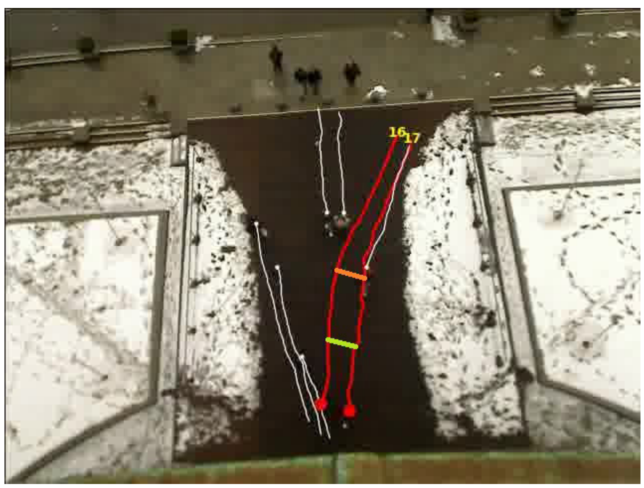

Figure 6: The group of two entites (red trajectories) are moving towards the south. The orange and light green lines shows the end and the new start of the same group, respectively. Swarms can detect the group, while other definitions cannot because the length of contiguous time interval is below the required threshold δ.

Now, we look into details on the maximal groups found by each definition. It turns out that all maximal groups found by the three other definitions are the same, while the swarm also finds them and more. From the ETH data set, we give an example of a maximal group that is found by swarms but not by the others (see Figure 6).

The example in Figure 6 is taken when we set δ to 89 or 105, while for $\delta = 72$, all definitions consider the two entities in red (number 16 and 17) as a maximal group. After 83 timestamps, they reach a position where their distance to each other exceeds ε (the orange line in Figure 6) and no other entities exist to keep their connectivity. Therefore, all grouping definitions that require a contiguous time interval as a duration of a group, fail to detect

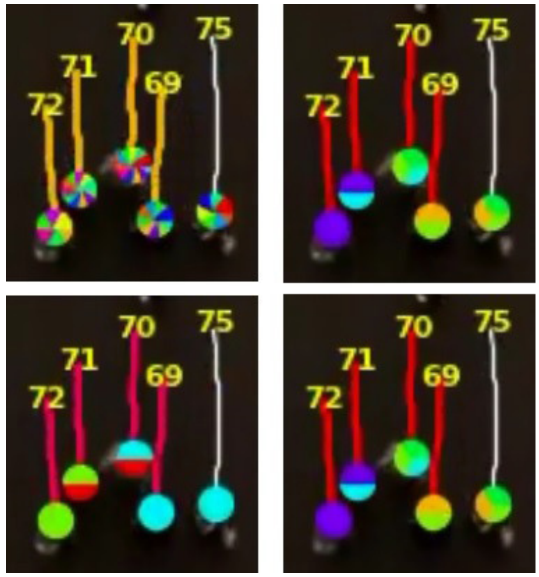

Figure 7: (Top-left) Swarm detects an annotated group of {69,70,71,72} but also many other groups compare to the other definitions. (Top-right) The result from convoy. (Bottom-right) The result from original group, which is the same as convoy. Notice that {70,75} is also a group. (Bottom-left) The result from refined group. Since their connectivity must be within their group only, {70,75} is not a group.

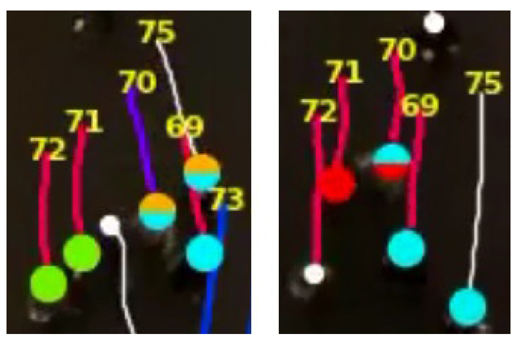

Figure 8: (Left) The white trajectory separates colored trajectories into two sub-groups. (Right) The leftmost pedestrian (72) moves faster and leaves the others.

the group for $\delta > 83$. Later, they resume their ε-closeness (starting at the light green line) for a short time, which enables swarms to detect this as a maximal group.

We present another example in Figure 7 where swarms (the top-left figure) find one particular maximal group (i.e., {69,70,71,72}) that other definitions cannot find. We show the reason in Figure 8. The left and right figures in Figure 8 show the events before and after the moment in Figure 7, respectively. In Figure 8 (left), the white pedestrian walks through the group and separates them far enough that the original group and convoy find the two sub-groups are disconnected. Although the two subgroups join again afterward, they do not stay together long enough since the leftmost pedestrian (72) moves faster and leaves the others, see Figure 8 (right). The time interval of these two events is not long enough for the required minimum duration of $\delta = 72$.

From Figure 7 (top-right) and (bottom-right), we can see that convoys and original groups have the same results, but the refined group (Figure 7 (bottom-left)) is different. Since the refined group only realizes connectivity within the members of groups, it does not consider the group of {70,75}. However, convoys and original groups use another entity (69) as an intermediate to keep the two pedestrians (70 and 75) to be ε-connected.

4.1.2 Vittorio Emanuele II Gallery.

The VEIIG data set is taken from the video surveillance in a hallway inside the Vittorio Emanuele II Gallery in Milan, Italy. The flow of entities in the video is mostly bidirectional. The results of our experiments are in Table 5.

		Precision	Recall	F1-score
$\delta = 17$	OG	0.199	0.952	0.329
	RG	0.223	0.947	0.361
	CO	0.202	0.952	0.333
	SW	0.125	0.957	0.221
$\delta = 38$	OG	0.357	0.884	0.509
	RG	0.414	0.884	0.564
	CO	0.362	0.884	0.514
	SW	0.238	0.932	0.379
$\delta = 58$	OG	0.444	0.778	0.565
	RG	0.503	0.778	0.611
	CO	0.451	0.778	0.571
	SW	0.315	0.870	0.463

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 5: Comparative results on the Vittorio Emanuele II data set.

First, we note that for all definitions, the precision values are relatively small. This can be explained by the fact that all definitions (except for swarm) consider a set of entities that is together during two disjoint but sufficiently long time intervals as two different (maximal) groups. Also, a group of 3 (or more) entities is often also found as one or two subgroups of 2 entities with a slightly longer duration. Therefore, we focus on the relative precision values. This also holds for the other data sets in our experiments.

We observe that in this data set, the refined group corresponds better to human annotation than the others based on their F1-score. This is mostly because the refined group definition has the highest precision out of the definitions considered. The swarm definition has the best recall value, while the maximal groups by the original group and convoy definitions find the same number of annotated groups; refined group misses one in total. The higher recall of swarm is related to the lower precision: swarm outputs many more groups, some of which correspond to human annotation. These may be groups with interrupted duration.

Our qualitative review shows several reasons why the grouping definitions cannot match all annotated groups, see Figure 9. One main reason is that the members of an

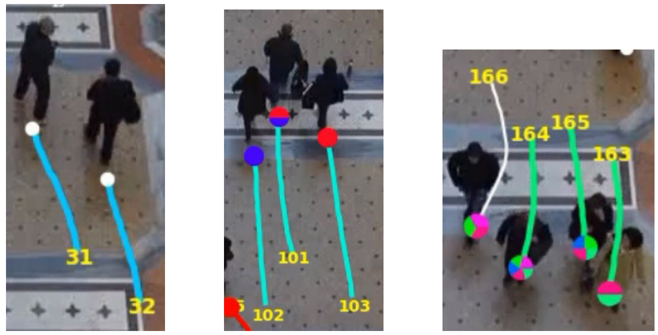

Figure 9: (Left) The pedestrians in a group are not within distance ε long enough. (Middle) The pedestrian in the middle is close to the left pedestrian and not close to the right one during this snapshot, but the reverse was the case earlier in the video. The annotation has all three in a group. (Right) The pedestrian on the left is always close to a group but the annotation does not include it.

Figure 10: (Left) In this group of 3, the two pedestrians on the right appear earlier and disappear later in the videos, making a maximal group of 2 that is a subgroup of the 3. (Right) Frames in sequence from left to right show a group that separated for a while, resulting in two different maximal groups by the grouping definitions, except for swarms.

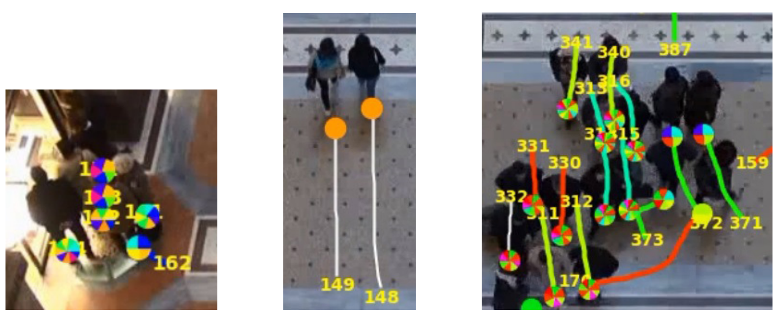

Figure 11: (Left) Two groups of pedestrians standing close together, making different maximal groups when they walk again. (Middle) A group found by all four definitions that was not annotated. (Right) In a dense environment, many more groups are produced by all grouping definitions.

annotated group are not within ε distance for a duration δ. This results in (i) annotated groups not recognized at all, or (ii) grouping definitions only found subgroups of annotated groups. There are also situations where entities are always within distance ε, but they were not annotated as a group. It possible to increase the recall by increasing ε, but the precision is likely to go down. Figures 10 and 11 show several scenarios that help to explain the low precision of all four definitions.

4.1.3 Crowds by Example.

The Crowds by Example (CBE) data set records pedestrian movement outside a university building. The flow of pedestrians is different than in the VEIIG data sets: pedestrians move in various directions with varying speed. In this data set, vertices are sampled once every 6 frames. For experiments using this data set, we set ε based on Proxemics Theory [14], rather than the theory by Solera et al. [40] which seems to suggest an unrealistically small value for ε (namely 0.41m). Instead, the maximum far phase for a personal distance between pairs of individuals from Proxemics Theory gives $\varepsilon = 1.22$m. We present the results in Table 6. Although swarms have the same discrete handling of the input, it performs better on recall because it appears there are groups with interrupted duration that are not found by the other definitions.

In dynamic crowds, we expect that entities from the same group will not be close to each other all the time, which is one reason why swarms can have a high recall score. Hence, we do the same experiment with different values for the parameter ε, to see how the other definitions that find maximal groups by a longest consecutive timestamp will perform. We set $\varepsilon = 1.52$ and present the results in Table 7. As expected, we see an increase in recall and decrease in precision for all definitions. However, the F1-score results from the swarm get worse and other definitions now perform better for all choices of minimum duration δ. Our qualitative evaluation of the Crowds by Example data shows similar situations as for the VEIIG data set.

		Precision	Recall	F1-score
$\delta = 36$	OG	0.172	0.565	0.264
	RG	0.181	0.565	0.274
	CO	0.167	0.600	0.261
	SW	0.176	0.652	0.277
$\delta = 57$	OG	0.243	0.461	0.318
	RG	0.254	0.452	0.325
	CO	0.245	0.470	0.332
	SW	0.269	0.574	0.366
$\delta = 78$	OG	0.307	0.339	0.322
	RG	0.315	0.339	0.327
	CO	0.291	0.357	0.321
	SW	0.326	0.522	0.401

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 6: Comparative results on the CBE data set; $\varepsilon = 1.22$.

		Precision	Recall	F1-score
$\delta = 36$	OG	0.131	0.817	0.226
	RG	0.138	0.817	0.236
	CO	0.130	0.835	0.225
	SW	0.097	0.861	0.174
$\delta = 57$	OG	0.180	0.722	0.288
	RG	0.203	0.722	0.317
	CO	0.182	0.713	0.290
	SW	0.135	0.800	0.231
$\delta = 78$	OG	0.224	0.609	0.328
	RG	0.260	0.591	0.361
	CO	0.228	0.617	0.333
	SW	0.164	0.757	0.270

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 7: Comparative results on the CBE data set; $\varepsilon = 1.52$.

Research question (1): correspondence to human annotation. Over all experiments, the refined groups have a slightly higher F1-score in correspondence to human annotation than the original groups and convoys definitions, but they are usually close. The higher F1-score is caused by a better precision. The swarms definition sometimes corresponds better and sometimes worse to human annotation. It appears to depend on the precise parameter settings. We also observe that human annotation is likely subjective.

4.2 Dependence on Density

In the following experiments, we investigate how the maximal groups produced by each definition are affected by the density of the environment. Therefore, for each data set, we consider grouping in situations of increasing density. Arguably, dense situations are especially difficult for identifying groups.

4.2.1 The Netlogo Flocking Model

We generated several data sets using an adapted version of the NetLogo Flocking model [45, 46]. In the adapted model the entities start to turn when they approach the border (instead of wrapping around), and there is a small random component in the new direction of the entities. This same model was used by Buchin et al. [5] to test the definition of original groups.

In all experiments, the size of the environment is fixed and set to 256×256 units. See Figure 12 for a general impression of the moving entities in these data sets. We consider different densities by varying the number of entities n to be 200, 300, or 400, and generate data sets with 500 time stamps each. For each generated data set, we compute all maximal groups for all four definitions, with a fixed $\delta = 10$ and $m = 10$, but using three values of ε, namely 4, 5, and 6. We chose to vary ε because this distance value is related to density. Each experiment is performed 10 times and the average and standard deviation are computed. The results of these experiments are shown in Table 8. There are no results for swarm when $\varepsilon = 6$ and $n = 400$ due to exponential running time (in n) of the algorithm to compute them.

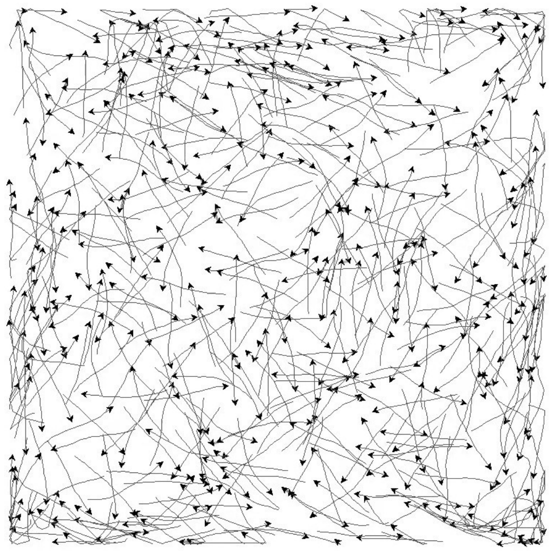

Figure 12: Trajectories from the Netlogo Flocking data set.

		average (10 sets)				std. dev.			
$\varepsilon = 4$	n	OG	RG	CO	SW	OG	RG	CO	SW
	200	0	0	0	0	0	0	0	0
	300	3.0	1.7	3.2	9.1	3.99	2.86	3.63	16.27
	400	23.1	12.3	31.1	112.2	23.48	17.95	21.65	252.86
$\varepsilon = 5$	n	OG	RG	CO	SW	OG	RG	CO	SW
	200	2.9	2.2	3.0	10.8	3.58	2.00	4.60	11.00
	300	41.6	38.4	61.7	229.0	12.17	6.50	12.79	77.89
	400	396.1	259.0	410.8	5299.6	64.15	47.61	53.61	2363.13
$\varepsilon = 6$	n	OG	RG	CO	SW	OG	RG	CO	SW
	200	33.9	25.9	36.6	229.3	9.63	7.98	9.83	45.00
	300	396.1	304.5	418.6	5017.3	108.00	64.43	106.38	2466.90
	400	1905.7	1357.0	1830.4	-	250.68	204.28	226.49	-

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 8: The average number of maximal groups for $m = 10$ and $\delta = 10$, and the standard deviation in the Netlogo data set for 10 generated sets.

For all definitions, the number of maximal groups increases as the density increases or when ε increases. Furthermore, the refined group produces fewer maximal groups than the other definitions, and swarm produces more. All definitions show a roughly 20-fold increase from $n = 200$ to $n = 300$ when $\varepsilon = 5$. From $n = 300$ to $n = 400$, the swarm definition has a larger than 20-fold increase, while the other three definitions have a less than 10-fold increase. For $\varepsilon = 4$, the values are too small for such observations. For $\varepsilon = 6$, we notice that the increase for swarm from $n = 200$ to $n = 300$ is much larger than for the other three definitions. Hence, it seems that swarm has a larger increase in the number of maximal groups than the other three definitions when the density or ε increases.

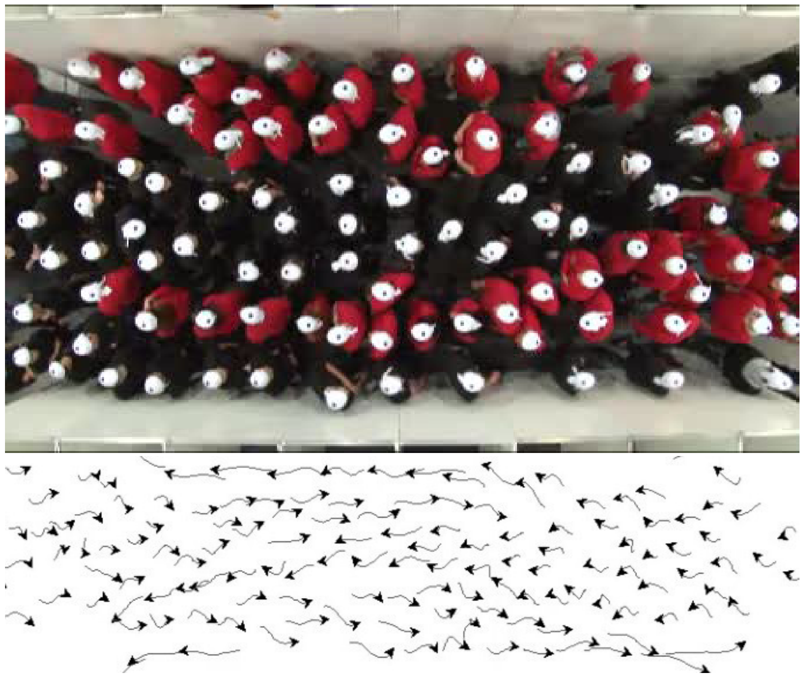

Figure 13: (Top) Frame from video recording of the corridor. (Bottom) Extracted trajectories from the video.

4.2.2 Pedestrians in a Synthetic Environment

The purpose of this experiment is to measure the conformity of a grouping definition in a dense environment. We use the data set which consists of trajectories extracted from video recordings of people walking in a synthetic environment [17, 53]. These trajectories are recorded by the Institute for Advanced Simulation of Jülich Supercomputing Centre to study the dynamics of pedestrians.

The particular data set we use consists of two sets of people walking in opposite directions through a corridor that is 8 meters long and 3.6 meters wide [22]. See Figure 13.[1] The density inside the corridor is controlled by the width w, in centimeters, of the two entrances to the corridor: a larger width w means that more people can enter the corridor simultaneously. The considered widths w are taken from $\{120, 160, 200, 250\}$. Each experiment consists of 300 trajectories, each of approximately 300 vertices as well.

In our experiments we fix the inter-entity distance ε to $80\,\text{cm}$, and choose the minimum group size m from $\{3, 6, 9\}$. For the minimum required duration δ we consider values in the range $[60, 180]$. This corresponds to a minimum group duration roughly between four and twelve seconds. For comparison, the average time $\bar{t}$ for a person to cross the corridor ranges from roughly twelve to twenty-three seconds. Note that since the density of the environment is extremely high, the swarm algorithm—whose running time is exponential—was unable to output any maximal group after a reasonable period of time (around 6 hours). Therefore, we exclude the swarm definition in these experiments.

[1]The frame is taken from http://ped.fz-juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/export/bot-360-250-250v2.mp4

	$m=3$				$m=6$				$m=9$			
	δ	OG	RG	CO	δ	OG	RG	CO	δ	OG	RG	CO
$w=120$, $\bar{t}=251.71$	60	901	341	886	60	691	177	682	60	582	116	579
	90	600	190	595	90	431	68	431	90	348	26	351
	120	394	117	394	120	261	31	264	120	200	7	203
	150	240	55	241	150	132	13	134	150	88	5	91
	180	144	35	141	180	59	7	55	180	40	3	37
	δ	OG	RG	CO	δ	OG	RG	CO	δ	OG	RG	CO
$w=160$, $\bar{t}=318.42$	60	4718	1126	4313	60	4416	867	4025	60	4200	747	3813
	90	3991	580	3669	90	3722	382	3414	90	3518	286	3215
	120	3371	387	3111	120	3118	232	2868	120	2923	154	2678
	150	2802	290	2595	150	2569	170	2370	150	2394	105	2204
	180	2246	229	2075	180	2037	138	1874	180	1871	85	1713
	δ	OG	RG	CO	δ	OG	RG	CO	δ	OG	RG	CO
$w=200$, $\bar{t}=369.23$	60	9749	2233	8018	60	9452	1959	7734	60	9194	1825	7488
	90	8700	1449	7164	90	8426	1232	6902	90	8180	1123	6669
	120	7697	959	6337	120	7431	768	6083	120	7198	672	5863
	150	6787	700	5616	150	6540	535	5378	150	6325	457	5168
	180	5948	538	4926	180	5722	397	4708	180	5510	338	4503
	δ	OG	RG	CO	δ	OG	RG	CO	δ	OG	RG	CO
$w=250$, $\bar{t}=374.38$	60	9277	2834	8396	60	9030	2611	8158	60	8777	2487	7921
	90	8205	1888	7438	90	7972	1693	7212	90	7725	1574	6981
	120	7280	1153	6617	120	7054	976	6400	120	6819	861	6182
	150	6406	680	5809	150	6190	524	5605	150	5963	422	5394
	180	5580	429	5064	180	5381	302	4876	180	5156	231	4667

Table 9: The number of maximal groups in the pedestrian data set.

The Number of Maximal Groups. The numbers of maximal groups for the considered parameter values are listed in Table 9. We first consider the number of maximal groups as a function of w, and thus of the density of the environment. As Figure 14 (left) highlights for the case $m = 6$ and $\delta = 150$, we see that up to $w = 200$, the number of reported maximal groups increase as a function of w. This applies for the three definitions of a group, although the number of maximal groups according to the original group and the convoy increases much faster than for the refined group. For even bigger values of w, the number of maximal groups flattens off, or sometimes even decreases. These results are more apparent for larger values of δ. When the density becomes higher, the speed of pedestrians becomes slower and pedestrians that are far apart are more likely to form a group. These will make a maximal group becomes larger (in size) and much longer (in duration) and consequently, will decrease the total number of maximal groups.

The number of maximal groups reported by the refined group definition is generally much smaller than the number of maximal groups reported by the other two definitions. This is also clearly visible in Figure 14 (right), where we show the number of maximal groups, with $m = 6$, and $w = 200$, as a function of δ. The graphs for different settings of m and w are similar. Here, we also see that the number of maximal groups decreases as we increase the minimum required duration (which is to be expected).

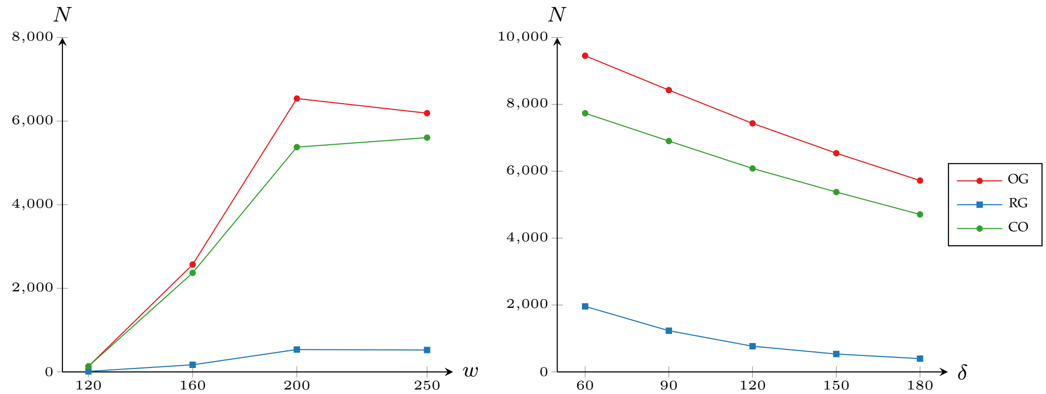

Figure 14: (left) The number of maximal groups (N) for $m = 6$ and $\delta = 150$ as a function of the width w of the corridor entrance, which influences density. (right) The number of maximal groups (N) for $m = 6$ and $w = 200$ as a function of δ. There are much fewer maximal groups according to refined groups when compared with original groups and convoys.

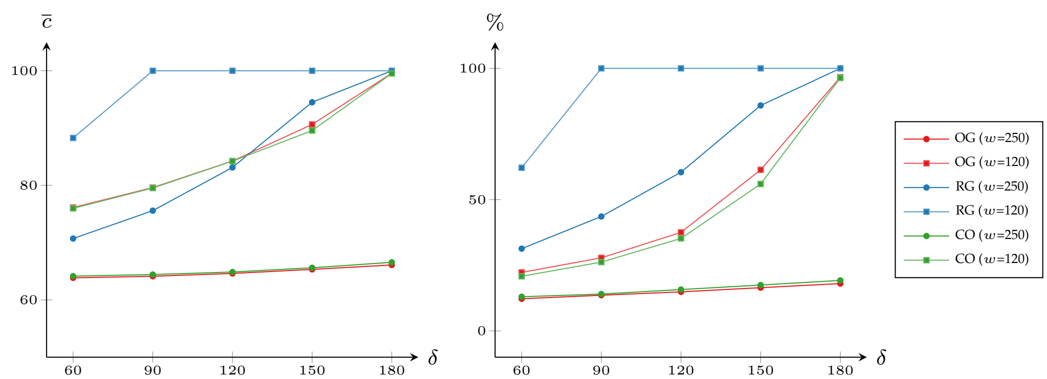

Figure 15: (left) The average number of conformity score ($\bar{c}$) in the pedestrian data for $m = 6$ and $w = 120$ as a function of δ. (right) The percentage of maximal groups with conformity 100 in the for $m = 6$ and $w = 120$ as a function of δ.

Measuring the Conformity of a Group. Since all entities (pedestrians) completely cross the corridor, we can classify each entity as type going "left to right" (type R), or "right to left" (type L). We can extend this notion to groups of entities by taking the type of the majority of its members (in case of ties we pick arbitrarily). We then define the *conformity* $c(\mathcal{G})$ of a group $\mathcal{G}$ as the percentage of its members that have the same type as the type of the group. Hence, the conformity of $\mathcal{G}$ is a value varying from 50, half of the members of $\mathcal{G}$ cross the corridor each way, to 100, all members of $\mathcal{G}$ go in the same direction. Intuitively, we expect that a set of people that act as a group (in the social sense) travel in the same direction, and thus we expect the conformity to be high in a good grouping definition.

We show the average conformity score in Figure 15 (left). We also consider the percentage of maximal groups that have conformity 100, that is, all group members travel in the same direction. We say that such a group is *uni-directional*. The results are in Figure 15 (right). Consider the case where $m = 6$ and $w = 120$. For all definitions, we see that as the minimum required duration increases, so does the average conformity score and the percentage of uni-directional maximal groups. However, the refined group definition generally has a much higher average conformity score and percentage of uni-directional maximal groups. In particular, for a duration as short as 90 time units (about 5 seconds), all maximal groups are uni-directional. For the original group and the convoy, this requires a minimum duration threshold of more than 180. These results are even more clearly visible as we increase the width of the corridor. For example, for $w = 250$, all maximal groups from the refined group definition with a duration of at least $\delta = 180$ are uni-directional, whereas in the original group and the convoy definition, less than 20% of the reported maximal groups are uni-directional, even if we increase the minimum required duration to 180. We expect that this is mostly due to the fact that the original group and the convoy report many more maximal groups than the refined group.

Research question (2): dependence on density. As expected, all grouping definitions find more groups when the density of entities increases or when connectedness is satisfied at larger distances. The swarm definition has a larger increase in the number of maximal groups than the other definitions. Since we do not have human-annotated groups data, we cannot draw further conclusions from these observations.

However, in an extreme case where the density is very high and entities only move in opposite directions, the refined group appears to be more natural. The original group and convoy report many groups consisting of entities that move in opposite directions, whereas the refined group finds only a few of them. Another interesting observation is that the refined group gives fewer groups. We believe that the type of connectivity used by grouping definitions might be the reason behind this finding. The refined group considers connectivity which is much more restricted than the other definitions (refer to Table 1). It is not clear whether this is an advantage or a disadvantage since the nature of all definitions gives rise to groups that share entities at the same time.

4.3 Dependence on Sampling Rate

The purpose of our last experiment is to examine how different sampling rates of trajectories affect the maximal groups produced by each definition. We conduct experiments by gradually removing vertices from trajectories, thus decreasing their sampling rate. For each new data set consisting of trajectories with a lower sampling rate, we count how many maximal groups are found by each definition.

The data set consists of trajectories from pedestrians inside the Grand Central Terminal in New York City, USA (see Figure 16)[2]. The data set contains 6000 video frames in which data points are generated manually. This is once every 0.8 seconds. There are 12,684 pedestrians, with an average of 105.52 pedestrians in each frame. For our experiment, we choose two sets of 800 consecutive frames that have a high density. The first set contains 2591 trajectories while the second contains 3313 trajectories. The average number of vertices in a trajectory are 46.57 and 46.85, respectively.

[2]The background image and movement data are from [51].

Figure 16: The Grand Central Terminal and trajectories of pedestrians.

First, we create a homography matrix to map frame coordinates from the data set into ground coordinates. We choose $\varepsilon = 0.76$m for a personal distance between pairs, based on the maximum *close phase* from Proxemics Theory [14]. We vary the required minimum duration for a maximal group $\delta \in \{8, 12, 16\}$ in seconds. Finally, we consider different sampling rates for the two sets of trajectories by taking $25\%, 50\%, \text{and} 100\%$ of the vertices of the trajectories. Some trajectories may be removed because less than 2 vertices remain. The results of our experiments are in Tables 10 and 11.

We notice that the number of original groups and refined groups is stable or increases slightly when reducing the sampling rate. In contrast, the number of convoys decreases slightly and the number of swarms decreases substantially. This trend is related to the cumulative version of the time duration of swarms. Imagine a swarm with several disconnected time intervals on its time duration (with a sampling rate of 100%). By reducing the sampling rate, some of these time intervals are likely to disappear, or their length is reduced. Therefore, the swarm may not meet the required δ anymore. On the other hand, the other definitions which use consecutive timestamps are not affected much by this situation.

Research question (3): dependence on sampling rate. In general, it is preferable when a grouping definition is not influenced too much by the sampling rate, so in this respect the original and refined group definitions perform a bit better than convoys and much better than swarms.

5 Conclusions and Future Work

We experimentally evaluated four definitions for grouping in trajectory data: (ORIGINAL) GROUPS, REFINED GROUPS, CONVOYS, and SWARMS. We ran quantitative experiments to establish how well these definitions correspond to the human intuition of a group, how the

		25%	50%	100%
$\delta = 8s$	OG	174	178	170
	RG	173	177	169
	CO	140	158	177
	SW	153	180	249
$\delta = 12s$	OG	127	118	116
	RG	127	118	117
	CO	111	122	121
	SW	115	138	199
$\delta = 16s$	OG	97	96	96
	RG	97	97	96
	CO	92	94	97
	SW	98	111	162

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 10: The number of maximal groups from 2591 trajectories in the Grand Central Terminal data set with different sampling rate.

		25%	50%	100%
$\delta = 8s$	OG	276	268	257
	RG	269	262	255
	CO	222	256	264
	SW	229	259	379
$\delta = 12s$	OG	211	206	203
	RG	206	200	200
	CO	190	203	204
	SW	196	215	304
$\delta = 16s$	OG	172	168	159
	RG	167	163	157
	CO	160	156	161
	SW	166	183	252

OG = Original Groups, RG = Refined Groups, CO = Convoys, SW = Swarms

Table 11: The number of maximal groups from 3313 trajectories in the Grand Central Terminal data set with different sampling rate.

number of groups depends on the density of the entities in their environment, and how the number of groups depends on the sampling rate of the trajectories.

For qualitative assessment, we developed a style of video annotation that allows us to compare two different grouping definitions. It is best suited for comparisons to groups from human annotation. Videos using this visualization can be found on our website [47].

Conclusions. From our experiments, we draw three main conclusions.

1. **Recognizing human-annotated groups.** The REFINED GROUPS scores best in terms of recognizing all sets of entities that were a group according to the human annotations (highest F1-score). However, the difference is small: the ORIGINAL GROUPS, REFINED GROUPS, and CONVOYS definitions all perform similarly.

2. **Stability under different sample rates.** We observe that the definitions that consider the trajectories to be continuous mappings from time to space (ORIGINAL GROUPS and REFINED GROUPS) are more stable than the definitions considering the trajectories as discrete input (CONVOYS and SWARMS) when we consider the number of reported groups under reductions of the sampling rate.

3. **Difference between definitions**. In general, it appears that the SWARM definition is most different among the four definitions, suggesting that taking group duration consecutively or cumulatively has a larger effect on grouping than the discrete or continuous handling of the data, or the type of connectedness (see Table 1).

Future Work. We identify several directions for future research.

- Our first conclusion has a low confidence, since all results are similar. To be more conclusive in our experiments, we first of all need better human annotation (refer to Figure 11 (middle), similar cases also occur in other human-annotated data). In addition, we need more data sets. One interesting source is animals trajectory data sets (e.g., from Movebank [31]) since they may exhibit different behavior when grouping for various activities (e.g., foraging, migration) than pedestrians walking together.

- Our third conclusion raises the question what causes the SWARM definition to perform so differently. The definition is more robust to noise than the other methods because swarms count duration cumulatively rather than consecutively. On the flip side, the definition also finds more doubtful groups that arise from several short, by-chance encounters. Therefore, robust grouping definitions can be developed and compared, which would depend on a fourth parameter that describes how noise is handled. Examples are platoons [26] and robust groups [5]. Note that the extra parameter is expected to make proper experimentation harder.

- All four definitions considered in our experiments depend heavily on the spatial component to define the “togetherness" between entities (at certain times). However, the spatial component only concerns the distance between entities. Nevertheless, we can further explore other features to extend the approach to define togetherness so that it is not influenced solely by the distance between entities, for example, using measures for trajectories/groups of trajectories (e.g. density, formation stability, etc.) [49] or contextual information and semantic data of the trajectories (e.g., consider the type of entities and their relation such as parent-child or leader-follower relationship) [2,6].

Acknowledgments

M.v.K. and M.L. are partially supported by the Dutch Research Council (NWO) on the Commit2Data project “Geometric Algorithms for the Analysis and Visualization of Heterogeneous Spatio-temporal Data” (no. 628.011.005). M.L. is partially supported by the Dutch Research Council (NWO) on grant no. 614.001.504. F.S. is partially supported by the Dutch Research Council (NWO) on grant no. 612.001.651.

References

[1] ANDERSSON, M., GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica 12*, 4 (2008), 497–528. doi:10.1007/s10707-007-0037-9.

[2] ANDRIENKO, G., ANDRIENKO, N., AND HEURICH, M. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science 25*, 9 (2011), 1347–1370. doi:10.1080/13658816.2011.556120.

[3] BANDINI, S., GORRINI, A., AND VIZZARI, G. Towards an integrated approach to crowd analysis and crowd synthesis: A case study and first results. *Pattern Recognition Letters 44* (2014), 16–29. doi:10.1016/j.patrec.2013.10.003.

[4] BENKERT, M., GUDMUNDSSON, J., HÜBNER, F., AND WOLLE, T. Reporting flock patterns. *Computational Geometry 41*, 3 (2008), 111–125. doi: 10.1016/j.comgeo.2007.10.003.

[5] BUCHIN, K., BUCHIN, M., VAN KREVELD, M., SPECKMANN, B., AND STAALS, F. Trajectory grouping structure. *Journal of Computational Geometry 6*, 1 (2015), 75–98. doi:10.20382/jocg.v6i1a3.

[6] BUCHIN, M., DODGE, S., AND SPECKMANN, B. Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science 9*, 1 (2014), 101–124. doi:10.5311/JOSIS.2014.9.179.

[7] COLLES, F., CAIN, R., NICKSON, T., SMITH, A., ROBERTS, S., MAIDEN, M., LUNN, D., AND DAWKINS, M. S. Monitoring chicken flock behaviour provides early warning of infection by human pathogen Campylobacter. *Biological Sciences 283*, 1822 (2016), 20152323. doi:10.1098/rspb.2015.2323.

[8] DE LUCCA SIQUEIRA, F., AND BOGORNY, V. Discovering chasing behavior in moving object trajectories. *Transactions in GIS 15*, 5 (2011), 667–688. doi:10.1111/j.1467-9671.2011.01285.x.

[9] DOHERR, M., CARPENTER, T., WILSON, D., AND GARDNER, I. Evaluation of temporal and spatial clustering of horses with Corynebacterium pseudotuberculosis infection. *American Journal of Veterinary Research 60*, 3 (1999), 284–291.

[10] DU, B., LIU, C., ZHOU, W., HOU, Z., AND XIONG, H. Catch me if you can: Detecting pickpocket suspects from large-scale transit records. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), ACM, pp. 87–96. doi:10.1145/2939672.2939687.

[11] ESTER, M., KRIEGEL, H., SANDER, J., AND XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)* (1996), pp. 226–231. doi:10.5555/3001460.3001507.

[12] ETH ZÜRICH COMPUTER VISION LAB. ETHZ - computer vision lab: Datasets. https://data.vision.ee.ethz.ch/cvl/aem/ewap_dataset_full.tgz. Last accessed: August 1, 2021.

[13] GUDMUNDSSON, J., AND VAN KREVELD, M. Computing longest duration flocks in trajectory data. In *Proc. of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems* (2006), pp. 35–42. doi:10.1145/1183471.1183479.

[14] HALL, E. *The Hidden Dimension*. Anchor Books, 1992.

[15] HUANG, Y., CHEN, C., AND DONG, P. Modeling herds and their evolvements from trajectory data. In *Proc. of the 5th International Conference of Geographic Information Science, GIScience* (2008), pp. 90–105. doi:10.1007/978-3-540-87473-7_6.

[16] HWANG, S.-Y., LIU, Y.-H., CHIU, J.-K., AND LIM, E.-P. Mining mobile group patterns: A trajectory-based approach. In *Proc. of the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, PAKDD'05* (2005), pp. 713–718. doi:10.1007/11430919_82.

[17] INSTITUTE FOR ADVANCED SIMULATION (IAS), FORSCHUNGSZENTRUM JÜLICH. Forschungszentrum Jülich - data repositories]. https://fz-juelich.de/ias/ias-7/EN/Expertise/Data/_node.html. Last accessed: August 1, 2021.

[18] JACOBS, A., SUEUR, C., DENEUBOURG, J. L., AND PETIT, O. Social network influences decision making during collective movements in brown lemurs (Eulemur fulvus fulvus). *International Journal of Primatology 32*, 3 (2011), 721–736. doi:10.1007/s10764-011-9497-8.

[19] JEUNG, H., YIU, M. L., ZHOU, X., JENSEN, C., AND SHEN, H. T. Discovery of convoys in trajectory databases. *Proceedings of the VLDB Endowment 1*, 1 (2008), 1068–1080. doi:10.14778/1453856.1453971.

[20] JIN, C., CHEN, D., ZHU, F., AND WU, M. Detecting suspects by large-scale trajectory patterns in the city. *Mob. Inf. Syst. 2019* (2019), 1837594:1–1837594:11. doi:10.1155/2019/1837594.

[21] KALNIS, P., MAMOULIS, N., AND BAKIRAS, S. On discovering moving clusters in spatio-temporal data. In *Proc. of the Advances in Spatial and Temporal Databases, 9th International Symposium, SSTD 2005* (2005), pp. 364–381. doi:10.1007/11535331_21.

[22] KEIP, C., AND RIES, K. Dokumentation von versuchen zur personenstromdynamik. Project Hermes, Institute for Advanced Simulation (IAS), Forschungszentrum Jülich, 2009.

[23] LEI, T. X., SUKOR, N. S. A., RAHMAN, N. A., ROHANI, M. M., AND HASSAN, S. A. Pedestrian route choice of vertical facilities at KLCC underground train station. In *Sustainable Transportation Infrastructures: Series 2*, B. D. Daniel, M. M. Rohani, and N. A. Termida, Eds. Penerbit UTHM, 2019.

[24] LERNER, A., CHRYSANTHOU, Y., AND LISCHINSKI, D. Crowds by example. *Computer Graphics Forum 26*, 3 (2007), 655–664. doi:10.1111/j.1467-8659.2007.01089.x.

[25] LETTICH, F., ALVARES, L. O., BOGORNY, V., ORLANDO, S., RAFFAETÀ, A., AND SILVESTRI, C. Detecting avoidance behaviors between moving object trajectories. *Data & Knowledge Engineering 102* (2016), 22–41. doi:10.1016/j.datak.2015.12.003.

[26] LI, Y., BAILEY, J., AND KULIK, L. Efficient mining of platoon patterns in trajectory databases. *Data & Knowledge Engineering 100* (2015), 167–187. doi:10.1016/j.datak.2015.02.001.

[27] LI, Y., HAN, J., AND YANG, J. Clustering moving objects. In *Proc. of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2004), pp. 617–622. doi:10.1145/1014052.1014129.

[28] LI, Z., DING, B., HAN, J., AND KAYS, R. Swarm: Mining relaxed temporal moving object clusters. *Proceedings of the VLDB Endowment 3*, 1 (2010), 723–734. doi:10.14778/1920841.1920934.

[29] LOGLISCI, C. Using interactions and dynamics for mining groups of moving objects from trajectory data. *International Journal of Geographical Information Science 32*, 7 (2018), 1436–1468. doi:10.1080/13658816.2017.1416473.

[30] MARX, M. Python 2d/3d trajectory visualization library. https://github.com/marximus/trackviz/. Last accessed: August 1, 2021.

[31] MAX PLANCK INSTITUTE FOR ANIMAL BEHAVIOR AND NORTH CAROLINA MUSEUM OF NATURAL SCIENCES AND UNIVERSITY OF KONSTANZ. Movebank for animal tracking data. https://www.movebank.org. Last accessed: August 1, 2021.

[32] MILLER, H., DODGE, S., MILLER, J., AND BOHRER, G. Towards an integrated science of movement: Converging research on animal movement ecology and human mobility science. *International Journal of Geographical Information Science 33*, 5 (2019), 855–876. doi:10.1080/13658816.2018.1564317.

[33] NASERIAN, E., WANG, X., XU, X., AND DONG, Y. A framework of loose travelling companion discovery from human trajectories. *IEEE Trans. Mob. Comput. 17*, 11 (2018), 2497–2511. doi:10.1109/TMC.2018.2813369.

[34] PELLEGRINI, S., ESS, A., SCHINDLER, K., AND GOOL, L. V. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proc. of the 12th IEEE International Conference on Computer Vision, ICCV 2009* (2009), pp. 261–268. doi:10.1109/ICCV.2009.5459260.

[35] RACHURI, K., MUSOLESI, M., MASCOLO, C., RENTFROW, P., LONGWORTH, C., AND AUCINAS, A. Emotionsense: A mobile phones based adaptive platform for experimental social psychology research. In *Proc. of the 12th ACM International Conference on Ubiquitous Computing* (2010), pp. 281–290. doi:10.1145/1864349.1864393.

[36] SAWAS, A., ABUOLAIM, A., AFIFI, M., AND PAPAGELIS, M. Tensor methods for group pattern discovery of pedestrian trajectories. In *19th IEEE International Conference on Mobile Data Management, MDM 2018* (2018), IEEE Computer Society, pp. 76–85. doi:10.1109/MDM.2018.00024.

[37] SHEIN, T. T., PUNTHEERANURAK, S., AND IMAMURA, M. Discovery of evolving companion from trajectory data streams. *Knowledge and Information Systems* (2020). doi:10.1007/s10115-020-01471-2.

[38] SHEN, J., CAO, J., LIU, X., AND TANG, S. SNOW: detecting shopping groups using wifi. *IEEE Internet Things J. 5*, 5 (2018), 3908–3917. doi:10.1109/JIOT.2018.2839525.

[39] SOLERA, F. Group detection and crowd analysis. https://aimagelab.ing.unimore.it/imagelab/researchActivity.asp?idActivity=08. Last accessed: August 1, 2021.

[40] SOLERA, F., CALDERARA, S., AND CUCCHIARA, R. Socially constrained structural learning for groups detection in crowd. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38*, 5 (2016), 995–1008. doi:10.1109/TPAMI.2015.2470658.

[41] TANG, L.-A., ZHENG, Y., YUAN, J., HAN, J., LEUNG, A., HUNG, C.-C., AND PENG, W.-C. On discovery of traveling companions from streaming trajectories. In *Proc. of the 2012 IEEE 28th International Conference on Data Engineering (ICDE '12)* (2012), pp. 186–197. doi:10.1109/ICDE.2012.33.

[42] TANG, L. A., ZHENG, Y., YUAN, J., HAN, J., LEUNG, A., PENG, W., AND PORTA, T. L. A framework of traveling companion discovery on trajectory data streams. *ACM Transactions on Intelligent Systems and Technology 5*, 1 (2013), 3:1–3:34.

[43] VAN KREVELD, M., LÖFFLER, M., STAALS, F., AND WIRATMA, L. A refined definition for groups of moving entities and its computation. *International Journal of Computational Geometry & Applications 28*, 2 (2018), 181–196. doi:10.1142/S0218195918600051.

[44] VIEIRA, M., BAKALOV, P., AND TSOTRAS, V. On-line discovery of flock patterns in spatio-temporal data. In *Proc. of the 17th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2009* (2009), pp. 286–295. doi:10.1145/1653771.1653812.

[45] WILENSKY, U. Netlogo. http://ccl.northwestern.edu/netlogo. Last accessed: August 1, 2021.

[46] WILENSKY, U., AND RAND, W. *An Introduction to Agent-based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. MIT Press, 2015.

[47] WIRATMA, L. Visualization of grouping definitions. https://tiny.cc/groupingvideos/. Last accessed: August 1, 2021.

[48] WIRATMA, L., LÖFFLER, M., AND STAALS, F. An experimental comparison of two definitions for groups of moving entities (short paper). In *Proc. of the 10th International Conference on Geographic Information Science, GIScience 2018* (2018), pp. 64:1–64:6. doi:10.4230/LIPIcs.GISCIENCE.2018.64.

[49] WIRATMA, L., VAN KREVELD, M., AND LÖFFLER, M. On measures for groups of trajectories. In *Societal Geo-innovation - Selected Papers of the 20th AGILE Conference on Geographic Information Science* (2017), pp. 311–330. doi:10.1007/978-3-319-56759-4_18.

[50] YAO, Q., SHI, Y., LI, H., WEN, J., XI, J., AND WANG, Q. Understanding the tourists' spatio-temporal behavior using open gps trajectory data: A case study of yuanmingyuan park (beijing, china). *Sustainability 13*, 1 (2021). doi:10.3390/su13010094.

[51] YI, S. Pedestrian walking path dataset. https://www.dropbox.com/s/7y90xsxq0l0yv8d/cvpr2015_pedestrianWalkingPathdataset.rar. Last accessed: August 1, 2021.

[52] YI, S., LI, H., AND WANG, X. Understanding pedestrian behaviors from stationary crowd groups. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015* (2015), pp. 3488–3496. doi:10.1109/CVPR.2015.7298971.

[53] ZHANG, J., KLINGSCH, W., SCHADSCHNEIDER, A., AND SEYFRIED, A. Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *Journal of Statistical Mechanics: Theory and Experiment 2012*, 02 (2012), P02002. doi:10.1088/1742-5468/2012/02/P02002.

[54] ZHAO, B., LIU, X., JIA, J., JI, G., TAN, S., AND YU, Z. A framework for group converging pattern mining using spatiotemporal trajectories. *GeoInformatica* (2020). doi:10.1007/s10707-020-00404-z.

[55] ZHENG, K., ZHENG, Y., YUAN, N. J., SHANG, S., AND ZHOU, X. Online discovery of gathering patterns over trajectories. *IEEE Transactions on Knowledge and Data Engineering 26*, 8 (2014), 1974–1988. doi:10.1109/TKDE.2013.160.