An Experimental Comparison of Two Definitions for Groups of Moving Entities

Lionov Wiratma[1]
Department of Information and Computing Sciences, Utrecht University
Utrecht, The Netherlands
l.wiratma@uu.nl
Department of Informatics, Parahyangan Catholic University
Bandung, Indonesia
lionov@unpar.ac.id

Maarten Löffler[2]
Department of Information and Computing Sciences, Utrecht University
Utrecht, The Netherlands
m.loffler@uu.nl

Frank Staals[3]
Department of Information and Computing Sciences, Utrecht University
Utrecht, The Netherlands
f.staals@uu.nl

Abstract

Two of the grouping definitions for trajectories that have been developed in recent years allow a continuous motion model and allow varying shape groups. One of these definitions was suggested as a refinement of the other. In this paper we perform an experimental comparison to highlight the differences in these two definitions on various data sets.

2012 ACM Subject Classification Computing methodologies → Model development and analysis, Theory of computation → Computational geometry

Keywords and phrases Trajectories, grouping algorithms, experimental comparison

Digital Object Identifier 10.4230/LIPIcs.GIScience.2018.64

Category Short Paper

1 Introduction

The presence of devices equipped with advanced tracking technologies, such as GPS-enabled mobile phones and RFID tags, makes it possible to easily record the position of moving entities over a period of time. The widespread use of such inexpensive devices leads to the availability of a vast amount of movement data. Consequently, in many research areas there is an increasing interest in analyzing such movement data [3, 11].

Typically, movement data is described as a *trajectory*: a path made by a moving entity over a period of time together with time stamps at the locations. Differently put, a trajectory is a continuous mapping from a time interval $I = [t_{start}, t_{end}]$ to the space in which the entity

[1] Supported by The Ministry of Research, Technology and Higher Education of Indonesia (No. 138.41/E4.4/2015)
[2] Supported by The Netherlands Organisation for Scientfic Research on grant no. 614.001.504
[3] Supported by The Netherlands Organisation for Scientfic Research on grant no. 612.001.651

10th International Conference on Geographic Information Science (GIScience 2018).
Editors: Stephan Winter, Amy Griffin, and Monika Sester; Article No. 64; pp. 64:1–64:6
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

is moving. An analysis task that has been well studied is to extract collective movement patterns from such data. Some of the movement patterns considered are flocks [1], herds [5], convoys [7], moving clusters [8], mobile groups [6] and swarms [9]. Buchin et al. formalize the definition for another variation called *groups* [2]. They define a group of moving entities by taking into account three parameters: the spatial parameter (are the entities close enough?), the temporal parameter (does the togetherness last long enough?), and the size parameter (are there enough entities?). They implement the algorithm to compute groups and present experimental evaluation of their method using both generated and real-world datasets. In a recent paper [10], we refined the definition of groups by Buchin et al. We made a slight change in the condition for the spatial parameter and argued that the refined definition of groups is more intuitive and is expected to be better for finding the right groups in a dense environment. Consequently, this change leads to different algorithms to compute groups.

In this paper we compare the two definitions experimentally. While there are many definitions of flocks, herds, groups, etc., the last two definitions and the flocking definition are the only ones that respect the continuity of the trajectories, and do not consider only fixed time-stamped locations. We exclude the flocking definition because it uses a fixed-size circle to define closeness, which does not allow for elongated groups. To compare the two grouping definitions, we implemented the algorithm to compute groups based on the refined definition (an implementation of the other one exists) and conducted experiments on dense pedestrian data. We compare the outputs from both implementations, which is the same as comparing the two definitions of groups, since the implementations follow the definitions exactly. We analyze the claim made (by us) in [10] that the newer definition is more intuitive, especially when the environment is dense. Arguably, dense situations are especially difficult for identifying groups.

Results and Organization. In the following section, we review both definitions, and highlight their differences. Section 3 briefly describes what we expect to find in an experimental analysis where we compare the two definitions. In Section 4, we describe our experiments. We focus our evaluation on the differences of the two definitions, and thus on the maximal groups that are reported, rather than the differences between the algorithms and their implementation. Moreover, we consider only a single dataset consisting of trajectories of pedestrians walking through a narrow corridor. We conclude in Section 5 where we discuss the advantages and disadvantages of the two definitions.

2 Description and Properties of the two Definitions for Groups

The original definition of a group by Buchin et al. relies on three parameters: the number of entities in a group, the time interval in which those entities form a group and the distance between entities in the group [2]. While the first two parameters are simple to formalize, the latter needs to be described in more detail. The ε-connectivity between two entities is defined as follows: Let $\mathcal{X}$ be a set of moving entities and consider two entities $x, y \in \mathcal{X}$. If at some time t, the Euclidean distance between x and y is at most ε ($\varepsilon > 0$), then x and y are *directly ε-connected*. 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. Then, with the maximum entity inter-distance ε, a minimum number of m entities in a group and a minimum required duration of δ, a subset $G \subset \mathcal{X}$ is a *group* during time interval I, if the following three conditions hold [2]:

- G contains at least m entities.
- I has a duration at least δ.

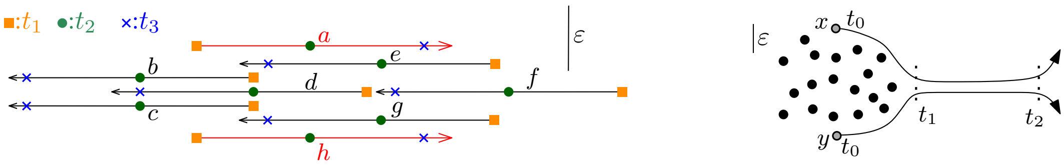

Figure 1 (left) Entities in $G = \{a, h\}$ are ε-connected using entities not in G [10]. (right) In the original definition [2], x and y are ε-connected during $[t_0, t_2]$.

- Every pair entities $x, y \in G$ is ε-connected in $\mathcal{X}$ during I.

Furthermore, G is a *maximal group* during time interval I if there is no time interval $I' \supset I$ for which G is also a group and there is no $G' \supset G$ that is also a group during I.

However, this definition might have a counter-intuitive effect and may not be suitable in a dense environment. In [10], we presented an example where this definition will have two entities in one group that are far apart during their entire duration as a group, see Figure 1 (left). Here, a and h are always ε-connected through different entities between t_1 and t_3. Hence, $\{a, h\}$ form a group during the time interval $[t_1, t_3]$. Since there is no superset of $\{a, h\}$ in the same time interval I, $\{a, h\}$ is a maximal group. Intuitively, we do not view $\{a, h\}$ as a group because they are separated by other entities that move in the opposite direction. To avoid this counter-intuitive situation, we refined the definition of a group by changing the requirement on the connectivity between entities in a group:

- Every pair entities $x, y \in G$ is ε-connected in G during I.

The only difference is that connectivity must happen using entities in the group G itself, and it can no longer use any entity in the whole set that is not part of the group. With this refined definition, $\{a, h\}$ is not a group because they are not ε-connected through other entities in the same group. Another example that shows the difference between the two definitions can be seen in Figure 1 (right) [10]. With the original definition, x and y are a group starting at t_0 because they are ε-connected through black entities that are standing still. However, by the refined definition, the group of $\{x, y\}$ starts only at t_1 when a and h encounter each other.

We compute all maximal groups according to the original definition using the algorithm of Buchin et al. [2]. For a set of n entities each specified using τ time-stamped locations, this algorithm runs in $O(\tau n^3 + N)$ time, where N is the output size. We use their original implementation. Computing all maximal groups according to refined definition [10] takes $O(\tau^2 n^5 \log n)$ time. We implemented the algorithm ourselves.

3 Expectations

The two definitions for groups differ only in a subtle way. We observe that any group by the refined definition is a group by the original definition, in particular, any maximal group by the refined definition is a (not necessarily maximal) group by the original definition. This implies that for any maximal group by the refined definition, there exists a maximal group by the original definition that has at least these entities and at least this duration.

We can expect that in situations that are "easy" for detecting groups, the two definitions give similar results in terms of the number of maximal groups and the duration of these groups. When the situation gets more and more complex, the detection of groups also gets more difficult. The small difference in the definitions may lead to different results now, because the accidental linking of entities through ε-closeness via entities that are not in the group is more likely to happen, which is exactly where the definitions differ. So we

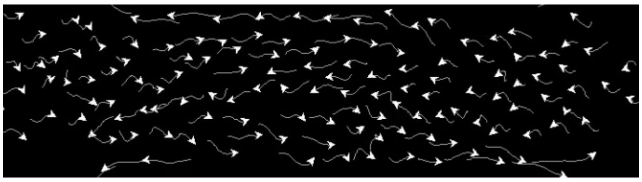

Figure 2 Trajectories of people walking in the corridor from the pedestrian data provided by the Jülich Supercomputing Centre.

may see maximal groups in the original definition that do not exist in the refined definition. Furthermore, maximal groups may have a longer duration by accidental linking just before the group is ε-connected or just after it.

It is not directly clear, however, that the original definition will return more maximal groups. Besides the effect just sketched above, it can also be that a maximal group in the original definition is briefly spread too much but some other entity in the neighborhood provides the linking to keep on seeing it as one maximal group. This linking would not be realized in the refined definition, which may lead to two maximal groups due to the interruption. If this happens much, the refined definition might give more maximal groups.

4 The Pedestrian Data

Our set of experiments uses pedestrian data collected by the Civil Security and Traffic division of the Jülich Supercomputing Centre [4] to study the dynamics of pedestrians. The data consists of trajectories extracted from video recordings of people walking in a synthetic environment. The particular datasets we use consist of two sets of people walking in opposite directions through a corridor that is 8 meters long and 3.6 meters wide [4]. 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 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 $\bar{t}$ it takes a person to cross the corridor ranges from roughly twelve to twenty-three seconds.

The Number of Maximal Groups. We first consider the number of maximal groups as a function of w, and thus of the density of the environment. As Figure 3 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 both the definitions of a group, although the number of maximal groups according to the original definition increases much faster than for the refined definition. 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.

The number of maximal groups reported by the refined definition is generally much smaller than the number of maximal groups reported by the original definition. This is also clearly visible in Figure 4, 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).

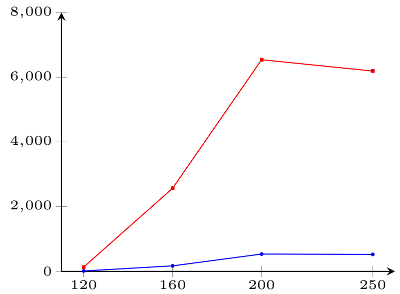

Figure 3 The number of maximal groups for $m = 6$ and $\delta = 150$ as a function of the width w of the corridor entrance, which influences density.

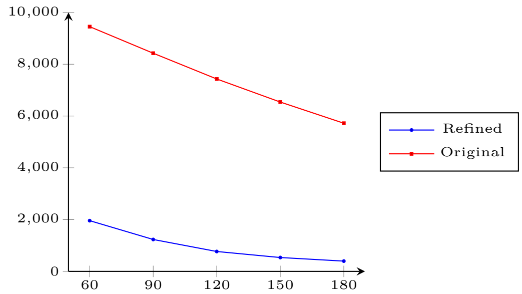

Figure 4 The number of maximal groups for $m = 6$ and $w = 200$ as a function of δ. There are much fewer maximal groups according to the refined definition when compared with the original definition.

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* $c(G)$ of a group G as the percentage of its members that have the same type as the type of the group. Hence, the conformity of G is a value varying from 50, half of the members of G cross the corridor each way, to 100, all members of 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 now measure the conformity of all maximal groups reported by our two definitions. Specifically, we 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 5. Consider the case where $m = 3$ and $w = 120$. For both definitions, we see that as the minimum required duration increases, so does the percentage of uni-directional maximal groups. However, the refined definition generally has a much higher 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 definition 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 $w = 160$, all maximal groups with a duration of at least $\delta = 120$ are uni-directional, whereas in the original definition less than 40% 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 definition reports many more maximal groups than the refined definition. We get similar results for larger minimum group size thresholds, that is, $m = 6$ and $m = 9$.

5 Conclusions

We examined two definitions for groups in trajectory data which both support continuous movement and varying shapes of groups. One definition was introduced as a refinement of the other, to obtain a more natural formalization of groups, but at the expense of a less efficient algorithm for their computation. Our comparison is based on a number of experiments where groups are computed by both definitions.

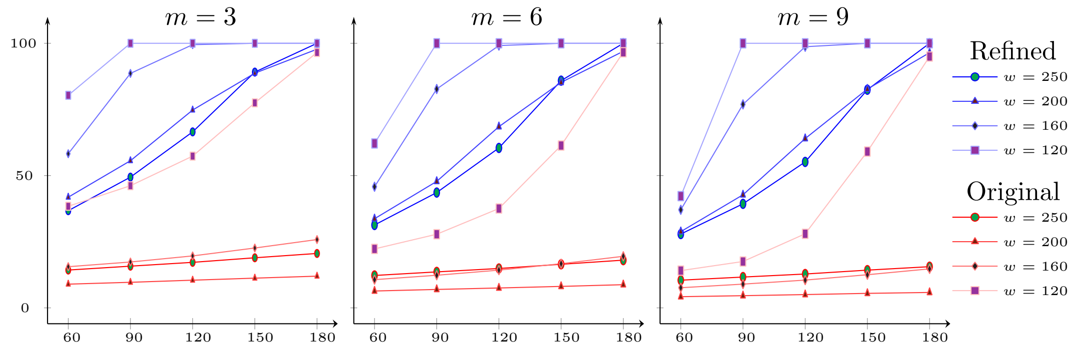

Figure 5 The conformity of the maximal groups in the pedestrian data as a function of δ.

The most important finding is that the two definitions differ more and more as the density of the crowd increases. This implies that in dense situations it does matter which definition is taken, even though they seem very similar. A second observation is that the refined definition appears to be more natural, at least in some cases. The original definition reports many groups that contain entities that move in opposite directions, whereas the refined definition finds only a few of them. Moreover, such groups then often have a short duration. Another interesting observation is that the refined definition gives fewer groups. It is not clear whether this is an advantage or a disadvantage, since the nature of both definitions gives rise to groups that share entities at the same time.

References

1 Marc Benkert, Joachim Gudmundsson, Florian Hübner, and Thomas Wolle. Reporting flock patterns. *Comput. Geom.*, 41(3):111–125, 2008.

2 Kevin Buchin, Maike Buchin, Marc van Kreveld, Bettina Speckmann, and Frank Staals. Trajectory grouping structure. *Journal of Computational Geometry*, 6(1):75–98, 2015.

3 Joachim Gudmundsson, Patrick Laube, and Thomas Wolle. Computational movement analysis. In Wolfgang Kresse and David N. Danko, editors, *Handbook of Geographic Information*, pages 725–741. Springer, Berlin, 2012.

4 Stefan Holl. *Methoden für die Bemessung der Leistungsfähigkeit multidirektional genutzter Fußverkehrsanlagen*. Dr., Bergische Universität Wuppertal, Jülich, 2016. Bergische Universität Wuppertal, Diss., 2016.

5 Yan Huang, Cai Chen, and Pinliang Dong. *Modeling Herds and Their Evolvements from Trajectory Data*, pages 90–105. Springer, Berlin, 2008.

6 San-Yih Hwang, Ying-Han Liu, Jeng-Kuen Chiu, and Ee-Peng Lim. Mining mobile group patterns: A trajectory-based approach. In *Proc. 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, PAKDD'05, pages 713–718. Springer, 2005.

7 Hoyoung Jeung, Man Lung Yiu, Xiaofang Zhou, Christian S. Jensen, and Heng Tao Shen. Discovery of convoys in trajectory databases. *PVLDB*, 1(1):1068–1080, 2008.

8 Panos Kalnis, Nikos Mamoulis, and Spiridon Bakiras. *On Discovering Moving Clusters in Spatio-temporal Data*, pages 364–381. Springer, Berlin, 2005.

9 Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays. Swarm: Mining relaxed temporal moving object clusters. *PVLDB*, 3(1):723–734, 2010.

10 Marc van Kreveld, Maarten Löffler, Frank Staals, and Lionov Wiratma. A Refined Definition for Groups of Moving Entities and its Computation. In *Proc. 27th International Symposium on Algorithms and Computation*, volume 64, pages 48:1–48:12. LIPIcs, 2016.

11 Yu Zheng and Xiaofang Zhou. *Computing with Spatial Trajectories*. Springer, 2011.