



## UvA-DARE (Digital Academic Repository)

### Intermittent and persistent movement patterns of dance event visitors in large sporting venues

Rutten, P.; Lees, M.H.; Klous, S.; Sloot, P.M.A.

**DOI**

[10.1016/j.physa.2020.125448](https://doi.org/10.1016/j.physa.2020.125448)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

Physica A: Statistical Mechanics and its Applications

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Rutten, P., Lees, M. H., Klous, S., & Sloot, P. M. A. (2021). Intermittent and persistent movement patterns of dance event visitors in large sporting venues. *Physica A: Statistical Mechanics and its Applications*, 563, [125448]. <https://doi.org/10.1016/j.physa.2020.125448>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



# Intermittent and persistent movement patterns of dance event visitors in large sporting venues

Philip Rutten<sup>a,\*</sup>, Michael H. Lees<sup>a</sup>, Sander Klous<sup>a</sup>, Peter M.A. Sloot<sup>a,b</sup>

<sup>a</sup> Informatics Institute, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands

<sup>b</sup> Institute for Advanced Studies, University of Amsterdam, 1012 GC, Amsterdam, The Netherlands

## ARTICLE INFO

### Article history:

Received 21 February 2020

Received in revised form 20 September 2020

Available online 22 October 2020

### Keywords:

Human mobility

Wi-Fi data

Positioning methods

Stretched exponential distribution

## ABSTRACT

There has been a number of reports showing evidence that human movement behaviour follows patterns resembling Lévy walks. These studies focus on the foraging patterns of rural humans and human hunter-gatherers. Here, we investigate motion patterns of visitors of a large dance event in the Johan Crujff Arena football stadium in Amsterdam. We find intermittent, persistent motion patterns. Using a path segmentation algorithm, we measure displacements (step lengths), and movement durations. We explore an alternative approach in the analysis of the movement tracks to overcome the limitations set by the bounded, concentric space of the building. Displacement distributions resulting from our alternative model deviate from the exponential and are best fit by a stretched exponential distribution. To further investigate the motion, we look at the mean-square displacement and autocorrelation of the turning angles. Although we find no evidence of Lévy walks, individuals move with directional persistence and superdiffusively up to a scale set by the size of the stadium.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The availability of data at increasingly fine spatial and temporal resolutions, together with a growing number of related research questions, has led to the emergence of the interdisciplinary research field of *movement ecology* [1]. The field of movement ecology has witnessed the development of several different random walk models for the description of animal motion. For many years, correlated random walk models have been a dominant conceptual framework for describing animal movement patterns [2]. In parallel there has been a growing number of reports showing that a wide variety of organisms exhibit movement patterns resembling Lévy flights or walks. This has given rise to the so-called Lévy flight foraging hypothesis, which states that movements of animals searching for various kinds of resources follow power law distributions (see [3] for an overview). Other studies have focused on the intermittency of animal movement, and have proposed models which combine alternating phases and scales [4–6].

Though not always from a foraging perspective, studies in movement behaviour have extended to humans too. Several pioneering studies in human mobility found that motion patterns of humans follow truncated power-law distributions [7,8]. Rhee et al. (2008) found truncated power-laws in GPS traces of volunteers in outdoor settings typically within a radius of tens of kilometers [9]. Raichlen et al. (2014) published evidence of Lévy walks in the movement patterns of 44 Hadza hunter-gatherers of northern Tanzania [10]. More recently, Reynolds et al. (2018) published evidence of Lévy walks in the movement patterns of rural humans in Mexico, and two different groups in Brazil [11]. Others, focusing on

\* Corresponding author.

E-mail address: [p.rutten@uva.nl](mailto:p.rutten@uva.nl) (P. Rutten).

human mobility on intra-urban scales, have found different distributions, such as log-normal [12], and exponential [13,14]. There is no agreement on which distribution best describes the empirical data (see [12] for an overview). Moreover, it is important to note that the validity of Lévy-like motion patterns in movement ecology is still controversial [15].

There is no previous research on human mobility that focuses exclusively on the movement of pedestrians, without interference of other transport means. Existing literature on human movement behaviour focuses on foraging patterns of rural humans and humans following a hunting and gathering lifestyle [10,11,16].

Here we analyse the movement behaviour of humans during an entertainment event, in a metropolitan context. We analyse movement data of visitors of a large dance event in the Johan Cruijff ArenA football stadium in Amsterdam. We use localisation of smart phones based on Wi-Fi detections to reconstruct individual trajectories, and use this as a proxy for human movement. The data provides insight into the movement dynamics of pedestrians that move freely in a large but bounded space, during more than 6 h of time. The bounded, concentric space of the Arena is not ideal for detecting large displacements. Due to the concentric layout, displacements are expected not to exceed the length of the stadium, as longer walks are forced into curved, circular trajectories. To overcome this limitation we explore an alternative approach in the analysis of the movement tracks, which allows for curvature in the trajectories. Our alternative approach allows us to measure arbitrarily large displacements, and facilitates possible underlying Lévy walk behaviour to emerge. This is a possibility we have to take into account, considering the existing evidence in animal and human movement behaviour.

The rest of the paper is organised as follows. In Section 2 we describe how we translate Wi-Fi measurements into a collection of movement tracks. We describe how we infer random walks from the movement tracks and measure displacements (step lengths), and movement durations. We introduce the statistical distributions included in the model selection. In Section 3 we describe the statistical results of our analysis, and discuss the findings in Section 4.

## 2. Methods

### 2.1. Data collection

We analyse data collected by the Wi-Fi network in the Johan Cruijff ArenA football stadium in Amsterdam. The wireless network consists of 591 access points (APs) with known spatial coordinates. The network is designed for complete coverage in the stadium. Though APs are not distributed homogeneously, AP locations are chosen to maximise coverage given the unique structure of the building and its accessible areas.

At a constant frequency (1 Hz) the APs switch to ‘monitor mode’ and capture all wireless traffic, regardless of destination addresses. The APs send reports of the monitoring results to a server where data are extracted, anonymised, and stored. The data contain the following information relevant for our research: the identity of the AP, the anonymised identity of source devices, received signal strengths (RSS) values, and timestamp indicating time of measurement.

We analyse Wi-Fi data collected during the Armin van Buuren dance event at May 12, 2017. Tickets for this event did not allocate seats, and people were free to walk around the stadium, including the pitch. The DJ stage was positioned along the northern edge of the playing field. A large outer corridor encircling the building was interrupted at that side of the building, being reserved for backstage functions. Other than that there were no obstacles present in the building, allowing the normal free flow of visitors throughout the stadium.

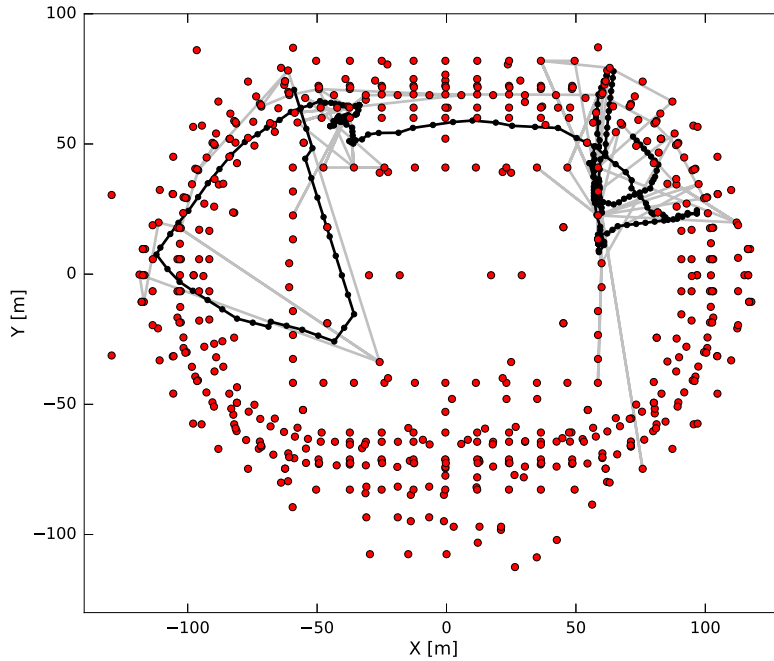
The data contains detections of 82,950 unique MAC addresses, collected during more than 9 h from 5:20 PM to 03:00 AM. When someone is not using his/her smart phone, the device eventually pauses all wireless communication. Therefore detection periods of devices alternate with periods without any detection. These time gaps range from small (seconds, minutes) to large (e.g. > 2 hours). We select data from devices with detection periods of minimum length  $\text{minperiod}$ , containing time gaps that do not exceed a maximum length  $\text{maxgap}$ . Here, we use ( $\text{minperiod} = 5, \text{maxgap} = 1$ ) minutes, which reduces the number of devices to 11,031 devices. So, for each device we have one or several detection periods ranging from 5 min to several hours, and spread over the more than 6 h of the event time.

We estimate locations of smart phones using proximity detection, which determines the position of a device based on its closeness to an AP with known spatial coordinates [17]. We store per time interval  $\Delta t = 10$  seconds the set of APs at which the device was detected, together with the strongest RSS value that occurred at each AP. We select the AP with the highest RSS value in the time interval. The time interval  $\Delta t = 10$  is chosen in order to maximise the number of detections within the interval, while minimising the amount of displacement that is ‘missed’ during one time step, assuming a pedestrian velocity of  $1 \text{ m}\cdot\text{s}^{-1}$ , and a spatial accuracy of order  $\sim 10$  meter.

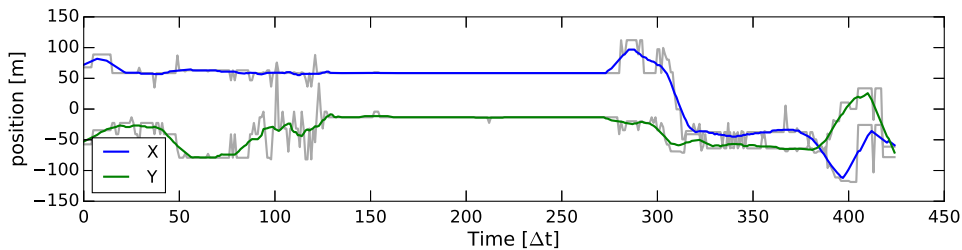
The proximity detection method produces temporal sequences of APs which are estimated to have been near a device. We ignore the z-coordinate and simplify the analysis to two dimensions. Although the accuracy of the proximity detection method is low, we use the positioning results for studying displacements only and do not pretend to accurately track individuals. Inaccuracies average out over the large numbers of displacements and are not expected to contribute substantially to displacement distributions. Also, the average distance between APs is on the order of 10 meters, while the displacement distribution extends from  $10^2$  to over  $10^3$  meters.

During crowded events, noise can reach levels that lead to considerable distortion of the distribution of RSS values over the APs. This is problematic for the proximity detection method, which simply selects the AP with the single highest RSS value in the time interval. As a result, AP sequences contain random fluctuation: jumps back and forth between APs that do not reflect real movement. To deal with this problem we use a simple moving average to smooth the movement tracks.

Drawing a line between successive smoothed location estimates produces a trajectory, or movement track (see Fig. 1).



**Fig. 1.** Example of a smoothed movement track (black line), together with the original AP sequence (grey line) produced by the proximity detection method. The original sequence contains some spurious long jumps. Tracks are characterised by sequences of jumps back and forth within a bounded region, alternated with sequences of jumps that constitute a larger displacement.



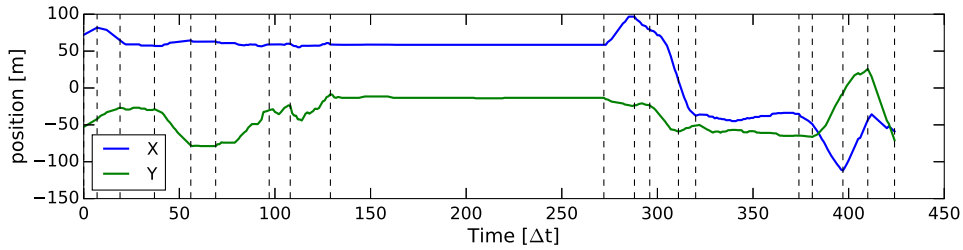
**Fig. 2.** Example of a track decomposed in two 1D time series (same as Fig. 1). Shown are the smoothed tracks (blue and green lines), together with the original sequences (grey lines) produced by the proximity detection method. The intermittent character of the movement data is clearly visible. Periods of rest alternate with periods of movement.

## 2.2. Movement track analysis

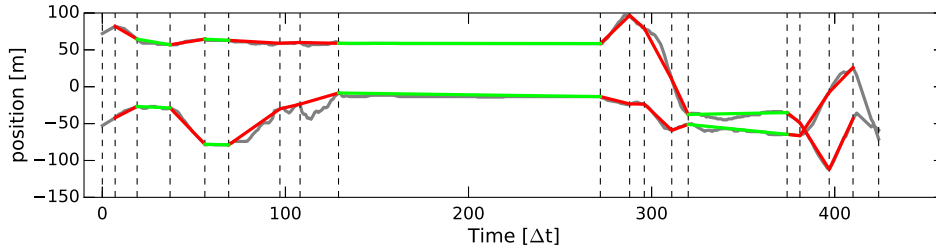
Movement tracks are characterised by (at least) two phases, or behavioural modes. Periods of rest alternate with periods of movement that form larger displacements. This aspect of the movement data shows up more clearly in 1D projections of the movement tracks onto the  $x$ - and  $y$ -axes (see Fig. 2). The observation also agrees with our intuition about human behaviour, as non-stop motion would be unlikely. People stay in one place for some time, and then decide to change location, usually in one continuous movement bout.

During the relocation phase individuals move continuously and with some degree of directional persistence. Due to the concentric layout of the building, movement is forced into curved, or even circular trajectories. To analyse the movement data within the random walk framework, we approximate it by series of straight line segments. From the segmented movement tracks we can measure step lengths, and movement durations. We use the Douglas–Peucker algorithm to segment the movement tracks [18]. The algorithm inserts break-points at large changes in the direction, and finds the change-points between movement and rest. We reconnect the break-points using straight lines. This method is very similar to the method introduced by Turchin (1998) [19] and used by Rhee et al. (2008) [9]. Below we describe the use of the Douglas–Peucker algorithm in more detail.

After the segmentation, we have moves of variable duration. Now several possibilities arise for defining a step length. A common approach is to set a critical turning angle  $\theta_c$  to select the turning points (e.g. [9,10]). However, as has been recognised, any choice of  $\theta_c$  is necessarily arbitrary [20]. Therefore we only explore two opposite extremes  $\theta_c = 0^\circ$  and



**Fig. 3.** Example of a track decomposed in two 1D time series (same as Fig. 1), together with the Douglas–Peucker segmentation. The dotted vertical lines represent the boundaries between the segments.



**Fig. 4.** Example of movement tracks with annotated waiting times (green), and displacements (red) (same track as in Figs. 1, 2, and 3). Displacements may consist of multiple segments connecting large turns in the movement track.

$\theta_c = 180^\circ$ , and compare the results. Note that these two variants are similar to the *rectangular* and *pause-based* models in [9]. When  $\theta_c = 0^\circ$ , a move is defined by one straight line segment. This approach is deemed not very interesting in our case, as observed step length distributions are expected to be truncated at the size of the stadium. When  $\theta_c = 180^\circ$ , a move consists of all the merged consecutive segments in between two periods of rest. The total length of the movement step is then given by merging all the segments. After merging we can reconnect start and end points of the whole movement episode as is done in [9]. However, this approach is again not fruitful in our case because we are in a circular arena. The resulting distribution is expected to be the same as when  $\theta_c = 0^\circ$ . Another possibility is to sum the sequence of lengths of straight line segments that make up a full movement episode in between two periods of rest. We feel this last approach most faithfully represents the actual movement behaviour. Step lengths correspond to actual behavioural events, instead of resulting from reorientations that are enforced by the layout of the building. Also, the step lengths more accurately describe the real distances travelled during the movement episodes.

In addition to the distribution of step lengths we study the movement durations  $t$ . Because pedestrians have finite velocity, larger displacements take more time to be completed. Therefore, movement duration can be used as a measure for displacement (as in [21–23]).

To break up the movement tracks in straight line segments we use the Douglas–Peucker algorithm [18]. First we decompose the 2D movement path into two 1D,  $x$  and  $y$ -coordinate time series (see Fig. 3). We apply the segmentation algorithm to the 1D time series in order to detect the pause times in a reliable way. For comparison, Rhee et al. use a threshold radius to determine whether lines between two consecutive locations are a ‘flight’ or a pause [9]. However, Wi-Fi data involve much larger positioning errors than GPS data, in which case the radius method produces many false classifications.

The Douglas–Peucker algorithm is a line simplification method. It recursively removes intermediate points from a polyline that are at a distance away from the line segment connecting first and last points, using a threshold parameter  $\epsilon$ . The value of the parameter  $\epsilon$  is chosen based on visual inspection of the results (see Appendix A for details). For each 1D time series, the algorithm produces a vector of change-points. To merge the two vectors we combine them in one time-ordered vector of unique change-points (as in [24]). To avoid over-segmentation, we iterate the combined vector and remove every change-point that follows the previous one in less than one minute. The resulting time granularity for the detection of behavioural episodes is 1 min, which agrees with our expectation, given the overall accuracy of the measurement system.

After the segmentation of the time series, we annotate each segment with the corresponding behavioural state (resting or moving). To do so, we apply a linear regression to each segment in both the 1D time series. Only if the slope parameters  $|\beta_x|$  and  $|\beta_y|$  of the two regression fits in both time series are below the threshold  $\alpha = 0.5$ , we label the segment as being stationary. Otherwise we label it as being part of a relocation move. The threshold value  $\alpha = 0.5$  is chosen based on visual inspection of the results (see Appendix A for details). In Fig. 4 we show an example of the annotated movement track, and its approximation by straight line segments.

### 2.3. Statistical model fitting

From the segmented movement tracks we measure step lengths  $\Delta r$ , and movement durations  $t$ . To determine which statistical model is underlying the movement process we first study the displacement distribution  $P(\Delta r)$ . People moving together in groups potentially threatens the statistical independence of the displacements. To test for underlying group structure, we apply a form of cluster analysis (see [Appendix D](#) for details). As we do not find compelling evidence of groups, no adjustments are made to the device selection. We fit four candidate distributions using the maximum likelihood estimation (MLE) methods of [22,25,26] (see [Appendix B](#) for more details). We determine whether the data are better fit by:

- The exponential distribution, with probability density function defined as:

$$p(x) = \lambda \exp(-\lambda(x - a)) \quad (1)$$

- The truncated Pareto distribution, with probability density function defined as:

$$p(x) = (1 - \mu)/(b^{1-\mu} - a^{1-\mu})x^{-\mu} \quad (2)$$

- The log-normal distribution, with probability density function defined as:

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) \quad (3)$$

- The stretched exponential distribution, with probability density function defined as:

$$p(x) = \beta\lambda x^{\beta-1} \exp(-\lambda(x^\beta - a^\beta)) \quad (4)$$

where  $a < x < b$ .

The exponential distribution is an indication of normal (Brownian) diffusion. While true Lévy walk behaviour leads to a full power law, the biologically constrained movement of humans within a bounded region can be reasonably identified as a truncated power law only. The log-normal distribution is a frequently reported model for describing heavy-tailed data [9,12]. The stretched exponential provides an alternative description of heavy-tailed data [27]. The stretched exponential provides the more convincing model, as its emergence is consistent with the random walk scheme describing the movement process (after merging and summing segments). This process is captured by a random walker that moves with a random velocity for some random amount of time, and for each move chooses its velocity and flight time from probability distribution functions  $h(v)$  and  $\phi(t)$ . Displacements are given by  $\Delta r = vt$ . Multiplicative processes, which involve the product of random variables, have been shown to give rise to stretched exponentials [27–29].

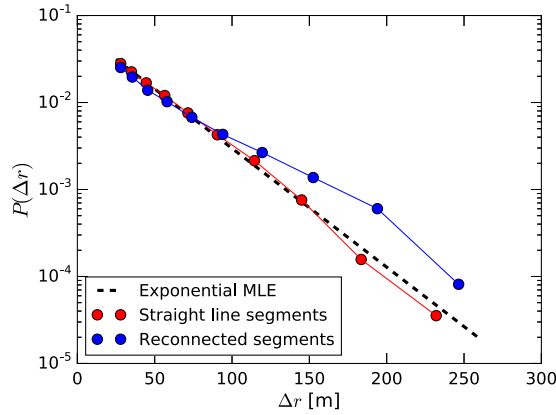
We also apply model fitting procedures to the movement durations  $\phi(t)$ . We test the same set of models, although the stretched exponential has not the same conceptual basis in the case of movement durations.

For both displacement distribution  $P(\Delta r)$  and movement durations  $\phi(t)$ , we select the most appropriate model to describe the data using the model selection method based on Akaike's information criterion (AIC) [30] (see [Appendix B](#) for details).

### 2.4. Movement characteristics

To further analyse the movement behaviour, we focus on the spatial characteristics of the movement tracks *during* the relocation moves. We test the observation that during the relocation moves individuals perform walks with directional persistence. In random walk models, persistence in direction is expressed through autocorrelation in turning angles between successive movement steps. This behaviour can be modelled using a correlated random walk (CRW) [19,31]. In concordance with the CRW approach we look at the statistical distribution of relative turning angles  $\varphi_i = \theta_i - \theta_{i-1}$ , where  $\theta_i$  is the direction of movement step  $i$ . Here, movement step refers to the step taken in one time step  $\Delta t$  introduced by the proximity detection method and used for the smoothed movement tracks (described in Section 2.1). Note that we thus measure the autocorrelation between steps *within* one relocation move, and not the correlation *between* different relocation moves.

Another important measure to test whether the relocation moves have directional persistence is to look at the diffusion behaviour. Diffusion is quantified by the mean squared displacement  $\langle(\Delta r)^2\rangle = \langle[r(t_0 + t) - r(t_0)]^2\rangle$  where the brackets  $\langle\dots\rangle$  denote averaging over all starting times  $t_0$ , and all movement tracks. CRWs lead to superdiffusion, i.e. diffusion for which  $\langle(\Delta r)^2\rangle \sim t^\gamma$ , with  $\gamma > 1$ , for time scales smaller than some characteristic correlation time  $\tau$ . On time scales larger than  $\tau$ , CRWs converge to normal (Brownian) diffusion ( $\gamma = 1$ ) [32]. Lévy walks on the other hand lead to genuine superdiffusion, i.e. diffusion for which  $\gamma > 1$ , on all time scales.



**Fig. 5.** Probability distribution function (PDF) of step lengths resulting from measuring separate straight line segments (red), on semi-log scale, together with the MLE fit of the exponential distribution, using  $a = 25$ . The data clearly follow a straight line on the semi-log scale, which suggests the displacements are well fit by the exponential distribution. Also shown is the distribution of step lengths after merging consecutive movement segments, and reconnecting with straight lines (blue). Both distributions are truncated at  $\sim 250$  m, as expected.

**Table 1**

Overview of the model selection results for the step lengths after merging and summing movement segments, based on Akaike weights, using  $a = 25$ .

Model	MLE parameters	95% CI	$\Delta$ AIC	Akaike weights
Exponential	$\lambda = 0.0114$	(0.0112, 0.0115)	2295	0
Truncated Pareto	$\mu = 1.7926$	(1.7916, 1.7936)	6213	0
Log-normal	$\mu = 4.3955$	(4.3868, 4.4041)	6345	0
	$\sigma = 0.7649$	(0.7588, 0.7711)		
Stretched exponential	$\lambda = 0.0816$	(0.0748, 0.0889)	0	1
	$\beta = 0.6561$	(0.6417, 0.6705)		

### 3. Results

We look at four different measures to characterise the movement: step lengths (displacements), movement durations, mean squared displacement (MSD), and turning angles. To determine which statistics are underlying the displacements we first look at the displacement distributions  $P(\Delta r)$ . We use lower cutoff value  $a = 25$ , which loosely defines the value after which the decay starts in the empirical probability distribution.

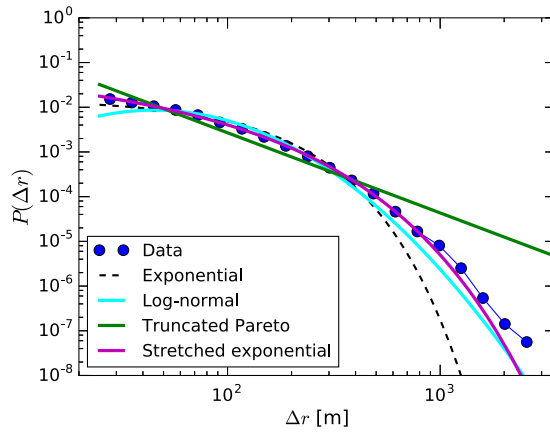
In Fig. 5 we show the step length distribution resulting from measuring separate straight line segments (in the  $\theta_c = 0^\circ$  approach) on semi-log scale. The data clearly follow a straight line on the semi-log scale, which suggests the displacements are exponentially distributed. We also show the maximum likelihood estimate (MLE) fit of the exponential distribution. In Fig. 5 we also show the step length distribution after merging consecutive movement segments, and reconnecting start and end points with straight lines. As expected the resulting distribution is similar, due to movements taking place in a circular arena. Displacements are truncated at  $\sim 250$  meter, which corresponds to the length size of the stadium.

In Fig. 6 we show the distribution of step lengths after merging consecutive movement segments and summing their lengths. We show the distribution on log-log scale, together with the MLE fits of the exponential, truncated Pareto, log-normal, and stretched exponential distributions. We select the most appropriate model using Akaike's information criterion (AIC). In Table 1 we show model selection results based on Akaike weights. The stretched exponential ( $w_{se} = 1$ ) provides the best description of the data. While the Akaike weights are decisively in favour of the stretched exponential, a visual inspection reveals that the difference in goodness-of-fit between the exponential and heavy-tailed distributions is subtle (see Fig. 6). A visual inspection persuades that the truncated Pareto distribution is a very poor fit to the data.

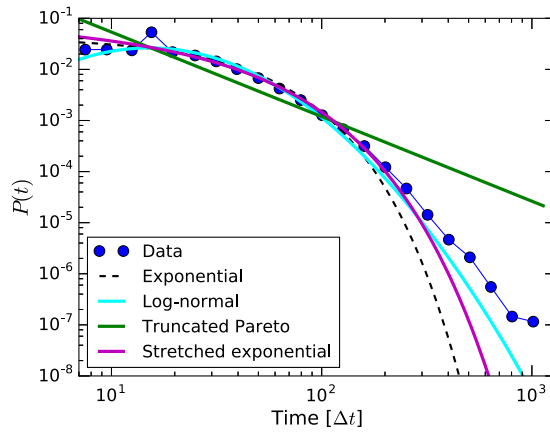
The MLE values of the stretched exponential are  $\lambda = 0.082$  and  $\beta = 0.66$ . The exponent value  $\beta = 0.66$  roughly corresponds to the theoretical value  $1/2$ , which is the inverse of the number of variables in the multiplicative process [27]. This is based on the hypothesis that displacement  $\Delta r = vt$  is the product of random variables  $v \sim h(v)$  and  $t \sim \phi(t)$ .

Next, we look at the movement duration distribution  $\phi(t)$ . In Fig. 7 we show the distribution of movement durations on log-log scale, together with the MLE fits of the exponential, truncated Pareto, log-normal, and stretched exponential. The stretched exponential is again the best fit, according to Akaike weights  $w_{se} = 1$ , and  $w = 0$  for all other models.

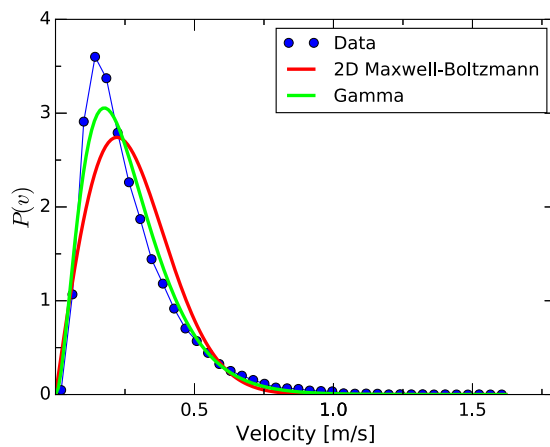
Measuring step lengths as well as movement durations allows us to evaluate average velocities during displacements  $v = \Delta r/t$ . Because the step lengths  $\Delta r$  may consist of multiple segments with different velocities,  $\Delta r/t$  corresponds to an average. In Fig. 8 we show the velocity distribution  $h(v)$  together with MLE fits of the Gamma and Rayleigh (2D Maxwell-Boltzmann) distributions. Model selection using Akaike weights shows that the Gamma distribution is the best fit ( $w_{\text{gamma}} = 1$  and  $w_{\text{rayleigh}} = 0$  with  $\Delta$ AIC = 4895). The Rayleigh (2D Maxwell-Boltzmann) distribution has



**Fig. 6.** Probability distribution function (PDF) of displacements after merging and summing line segments, on log–log scale, together with the MLE fits of the exponential, truncated Pareto, log-normal, and stretched exponential distributions, using  $a = 25$ . The Akaike model selection indicates that the stretched exponential provides the best description of the data (see Table 1).



**Fig. 7.** Probability distribution function (PDF) of movement durations on log–log scale, together with the MLE fits of the exponential, truncated Pareto, log-normal, and stretched exponential. Akaike model selection indicates that the stretched exponential provides the best fit.



**Fig. 8.** Probability distribution function (PDF) of the velocities. Also shown are the MLE fits of the Rayleigh (2D Maxwell–Boltzmann) distribution, and the Gamma distribution.

the more interesting implication, being the theoretical velocity distribution for gas particles in equilibrium (cf. [33]). The fact that the Rayleigh distribution does not provide the most accurate description for the data is not surprising.



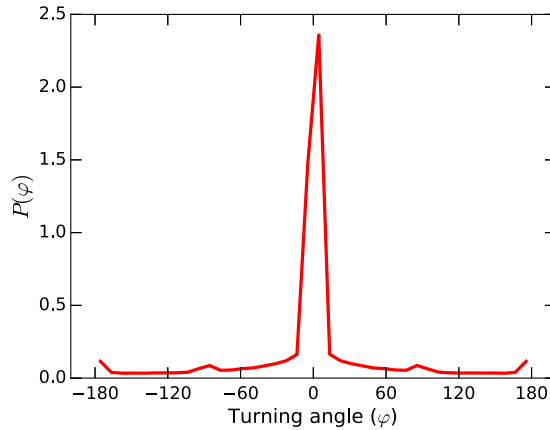


Fig. 9. Probability distribution of the turning angles.

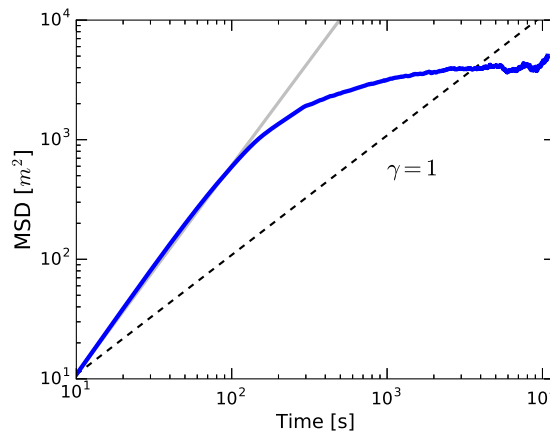


Fig. 10. Observed mean squared displacement (MSD) on log-log scale (blue line). The black dashed line corresponds to normal (Brownian) diffusion ( $\gamma = 1$ ). The grey line corresponds to superdiffusion ( $\gamma = 7/4$ ), and is shown as a visual aid.

The Rayleigh distribution arises when the velocity vector components in two dimensions are uncorrelated, zero mean, normally distributed random variables. In our case, velocities are calculated as the ratio of displacement and duration, which are both obtained following our alternative approach, using the path segmentation procedure.

The mean velocity according to the Gamma is  $\langle v \rangle = k/\beta = 0.265$ , where  $k$  and  $\beta$  are shape and rate parameters.

In Fig. 9 we show the probability distribution of turning angles. We see a sharp peak at zero which indicates the high degree of correlation in subsequent movement steps [19]. The autocorrelation in turning angles shows that individuals move with directional persistence *during* relocation moves. The movement tracks could thus be described as CRWs interrupted by rests.

In Fig. 10 we show the MSD  $\langle (\Delta r)^2 \rangle$ . Up to  $t \sim 10^2$  (seconds) the MSD approximates the line with slope  $\gamma = 7/4$  (solid grey), then decreases to  $\gamma < 1$  and shows truncation behaviour. The superdiffusion indicated by the slope in the first part of the MSD is consistent with displacements being described by CRWs. CRWs are known to give rise to superdiffusion up to a characteristic correlation time (e.g. [32], see also Appendix C). The transition of the MSD from values  $\gamma > 1$  to below 1 occurs somewhere in the region  $100 < t < 200$  seconds. This truncation time can be related to the time it takes to cross the stadium at moderate walking speed  $v \sim 1 \text{ m}\cdot\text{s}^{-1}$ . Obviously, directional persistence and superdiffusion cannot be maintained on scales larger than the size of the stadium.

Note that the time range in which truncation occurs in the MSD is much smaller than the time range of the measured displacements (as shown in Fig. 6). The measured displacements result from grouping time steps into relocation moves of variable duration, and are in the range  $70 < t < 11320$  seconds (see Fig. 7). This explains why the observed superdiffusion in the MSD does not affect the displacement distributions, as movement episodes are measured on time scales that stretch far beyond the region where the transition occurs from superdiffusion to truncation in the MSD.

## 4. Discussion

We have shown that motion patterns of dance event visitors are characterised by intermittency. Movement patterns are characterised by periods of rest, alternating with periods of movement. We have exploited this property of the motion by defining displacements as the distance travelled between two periods of rest. During the relocation phase people move with directional persistence, as evidenced by the turning angle analysis. The analysis of the MSD confirms that people move with directional persistence, as the MSD starts out in the superdiffusive regime until it shows truncation behaviour. Through the finite pedestrian velocity this truncation time can be roughly related to the size of the stadium.

We have shown that after merging and summing the segments in-between pause times, the range of step lengths greatly increases. We find that, after summing segments, heavy-tailed distributions provide better descriptions of the displacements than the exponential distribution. Visual inspection reveals however that the difference in goodness-of-fit between the different models is subtle. We find that the stretched exponential provides the more convincing model, as its emergence is consistent with the random walk scheme describing the movement process. This result is in agreement with the simple stochastic model proposed in Gallotti et al. (2016) [29] to describe various mobility patterns.

Despite our alternative approach to measure displacements, we find no evidence of Lévy walks. This is in contrast to several other studies of human movement behaviour [9–11]. A reasonable explanation is that the stadium does not invite displacements much larger than the size of the building, as this results in circular trajectories. This is in contrast with Rhee et al. who analyse GPS traces collected in outdoor settings typically on the scale of tens of kilometers [9]. The difference in results supports the idea that scale and structure of the environment influence the movement process that arises. For example, step length distributions may simply reflect spatial distances between various visited locations. This suggests that specific movement processes (such as Lévy walks) emerge from the interaction between the animal and environment, rather than from evolutionary adaptation of movement behaviour [3,34].

An important difference with the studies of Raichlen et al. (2014) [10] and Reynolds et al. (2018) [11] is that, in our case, we assume that dance event visitors are not (consciously) foraging. Also, there would be no clear benefit of walking very large distances during a dance event.

Despite the assumption that dance event visitors are not foraging we do find that displacements tend to become heavy-tailed, and that movement patterns have characteristics such as intermittence and persistence. All these characteristics have been found in many other animal movement studies and have been identified as optimal search strategies. These commonalities may point to more general mechanisms underlying different kinds of movement behaviour [35]. We conclude that clarifying these issues merits further research.

### CRedit authorship contribution statement

**Philip Rutten:** Conceptualization, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Michael H. Lees:** Supervision, Writing - review & editing. **Sander Klous:** Project administration, Funding acquisition, Writing - review & editing. **Peter M.A. Sloot:** Supervision, Writing - review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

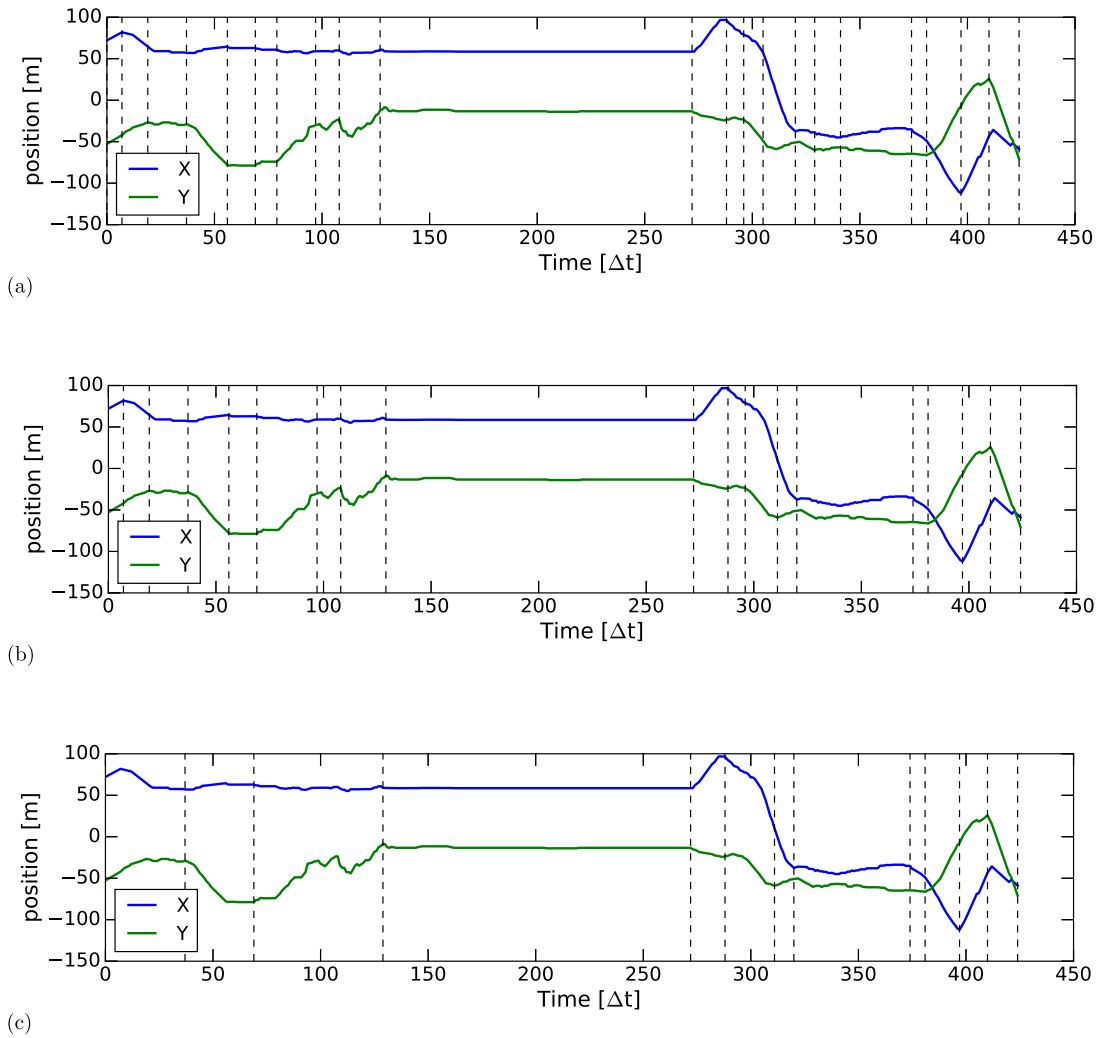
### Acknowledgement

We acknowledge the financial support of the Netherlands eScience Center, Netherlands, under grant number 027.015.001.

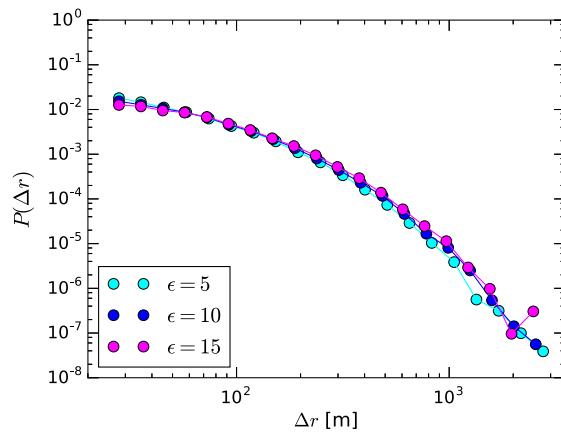
## Appendix A. Path segmentation and annotation

### A.1. Path segmentation

The Douglas–Peucker algorithm is a line simplification algorithm that uses a threshold parameter  $\epsilon$ . Here we illustrate the effect of different choices for the parameter value on the path segmentation. In Fig. 11(a–c) we show path segmentation results for the example movement track, for parameter values  $\epsilon = (5, 10, 15)$ . We find that  $\epsilon = 10$  produces the best result. In Fig. 12 we show the resulting probability distributions of the displacements, measured using our alternative approach. We see that the results are robust against variation in  $\epsilon$ . The stretched exponential distribution provides the best model for all three values of  $\epsilon$ , according to model selection based on Akaike weights. In Table 2 we show MLEs of the stretched exponential, for different values of  $\epsilon$ .



**Fig. 11.** Example track decomposed in two 1D time series, together with the Douglas–Peucker segmentation, using threshold parameter values (a)  $\epsilon = 5$ , (b)  $\epsilon = 10$ , (c)  $\epsilon = 15$ . The dotted vertical lines represent the boundaries between the segments.

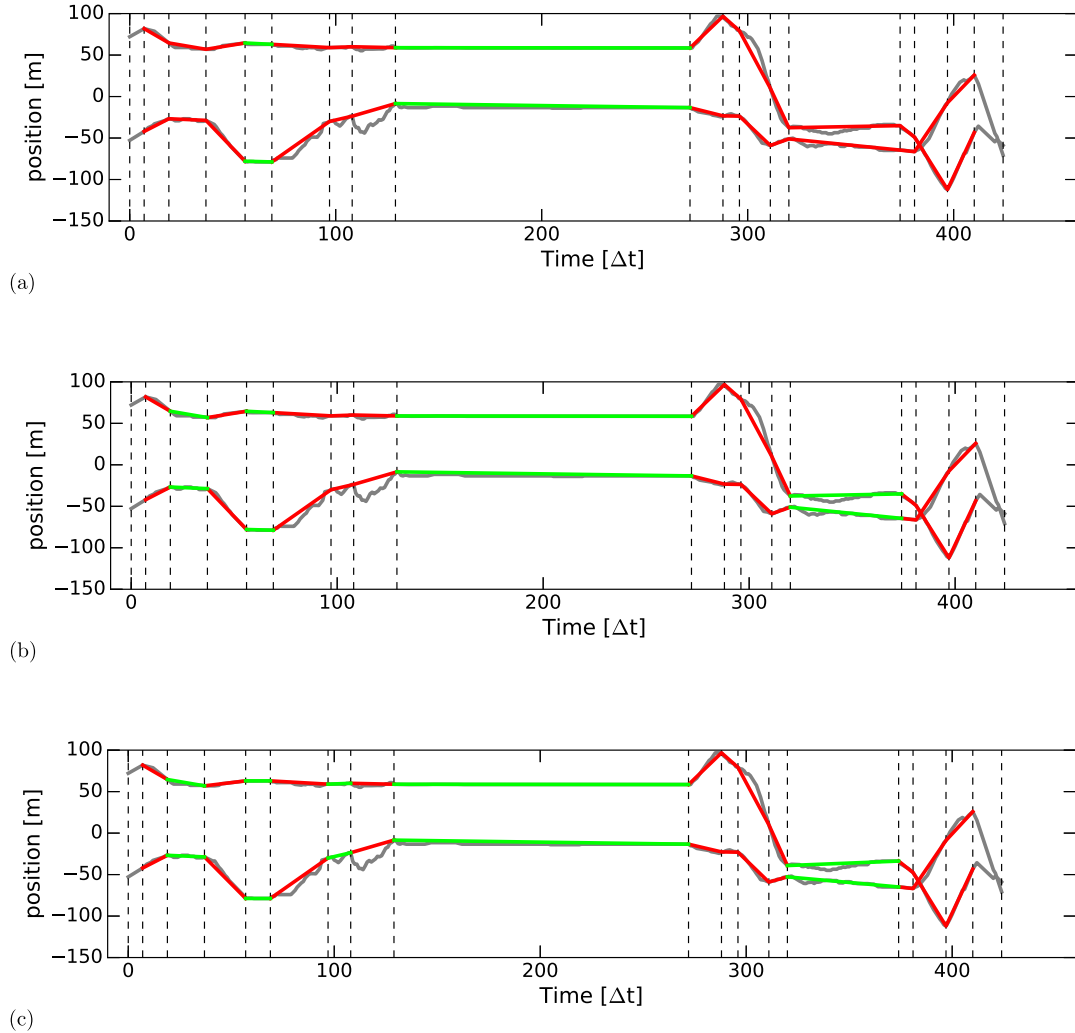


**Fig. 12.** Probability distribution function (PDF) of the displacements, for different values of  $\epsilon$ .

**Table 2**

Overview of fitting results using different values of  $\epsilon$ . Shown are the MLEs for the stretched exponential, which provided the best model for the data, according to model selection based on Akaike weights.

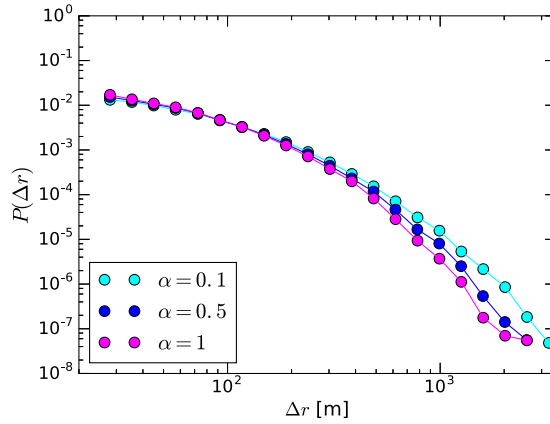
Parameter value	$(\Delta r)_{\max}$	$\lambda$	$\beta$
$\epsilon = 5$	3099.6	0.1089	0.6234
$\epsilon = 10$	4586.5	0.0816	0.6561
$\epsilon = 15$	4452.4	0.0591	0.6970



**Fig. 13.** Example track with annotated waiting times (green), and displacements (red), using threshold parameter values (a)  $\alpha = 0.1$ , (b)  $\alpha = 0.5$ , (c)  $\alpha = 1$ .

### A.2. Annotation of behavioural states

After the segmentation of the tracks, we annotate each segment with the corresponding behavioural state (resting or moving). To do so, we apply a linear regression to each segment in both the 1D time series, and use a threshold value for the slope  $\alpha = 0.5$ . Here we illustrate the effect of different choices for  $\alpha$  on the annotation result. In Fig. 13(a-c) we show path segmentation results for the example movement track, for parameter values  $\alpha = (0.1, 0.5, 1)$ . We find that  $\alpha = 0.5$  produces the best result. In Fig. 14 we show the resulting probability distributions of the displacements. We see that the results are robust against variation in  $\alpha$ . The stretched exponential distribution provides the best model for all three values of  $\alpha$ , according to model selection based on Akaike weights. In Table 3 we show MLEs of the stretched exponential, for different values of  $\alpha$ .



**Fig. 14.** Probability distribution function (PDF) of the displacements, for different values of  $\alpha$ .

**Table 3**

Overview of fitting results using different values of  $\alpha$ . Shown are the MLEs for the stretched exponential, which provided the best model for the data, according to model selection based on Akaike weights.

Parameter value	$(\Delta r)_{\max}$	$\lambda$	$\beta$
$\alpha = 0.1$	4599.7	0.0833	0.6317
$\alpha = 0.5$	4586.5	0.0816	0.6561
$\alpha = 1.0$	4599.7	0.0801	0.6751

## Appendix B. Statistical methods of model comparison

We use statistical methods of Clauset et al. [25] and Edwards et al. [22] for fitting the distributions. We benchmark results against the Python `powerlaw` package [26], which has implemented the methods from [25].

### B.1. Maximum likelihood estimation

The maximum likelihood estimate (MLE) of the parameter  $\lambda$  of the exponential distribution  $p(x) = \lambda e^{-\lambda(x-a)}$  is given by

$$\hat{\lambda} = 1 / \left( \sum_{i=1}^n x_i / n - a \right) \tag{5}$$

where  $n$  is the number of data points, and  $a$  is the lower bound of the fitting range. In this research  $a$  is loosely determined as the value after which the decay starts in the pdf.

There are no analytical solutions for the MLEs of the parameters in the log-normal distribution, and the stretched exponential distribution. In these cases we numerically minimise the negative log-likelihood function

$$L(\theta) = - \sum_{i=1}^n \log p(x_i | \theta) \tag{6}$$

where  $\theta$  is a vector of parameters in the model. The MLE of the parameter  $\mu$  in the truncated Pareto distribution is given by the numerical solution of the equation

$$\frac{1}{n} \sum_{i=1}^n \log x_i = \frac{1}{\mu - 1} + \frac{b^{1-\mu} \log b - a^{1-\mu} \log a}{b^{1-\mu} - a^{1-\mu}} \tag{7}$$

For both numerical minimisation and solution we use Python library functions (following [26]).

### B.2. Confidence intervals of estimated parameters

To compute confidence intervals we use the likelihood profile method [36]. The method compares the likelihood of the MLE of a parameter  $\theta$  with other values of that parameter. According to statistical theory

$$\mathcal{R} = 2[L(\theta) - L(\theta_{\text{MLE}})] \tag{8}$$

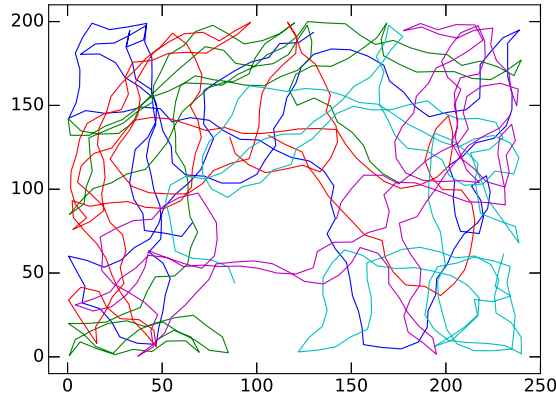


Fig. 15. Movement trajectories of 5 simulated correlated random walkers.

has a chi-square distribution with one degree of freedom. We can find 95% confidence interval boundaries by using the fact that  $\Pr\{\chi^2 < 3.84\} = 0.95$ , and numerically solving  $L(\theta) - L(\theta_{\text{mle}}) = 1.92$ . For models with two parameters we perform the test for each parameter separately. We systematically vary the parameter of interest, and at each instance compute the value for the other parameter that maximises the likelihood at that point.

### B.3. Akaike model selection

To compute Akaike weights we need the Akaike Information Criterion (AIC)

$$\text{AIC} = 2L(\theta_{\text{mle}}) + 2K \quad (9)$$

for which we require the value of the negative log-likelihood function at the maximum (MLE), and where  $K$  is the number of parameters to be estimated [22,30]. The AIC differences are

$$\Delta_i = \text{AIC}_i - \text{AIC}_{\min} \quad (10)$$

where  $\text{AIC}_{\min}$  is the AIC of the model with the minimum AIC, which is considered as the best model. The Akaike weights are given by

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{m=1}^M \exp(-\frac{1}{2}\Delta_m)} \quad (11)$$

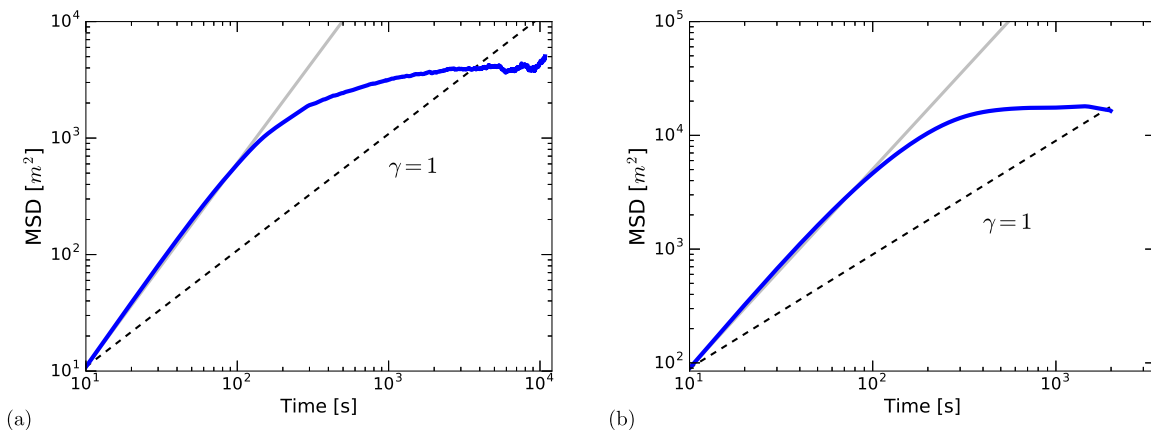
where  $M$  is the set of models to be compared.

## Appendix C. Mean square displacement

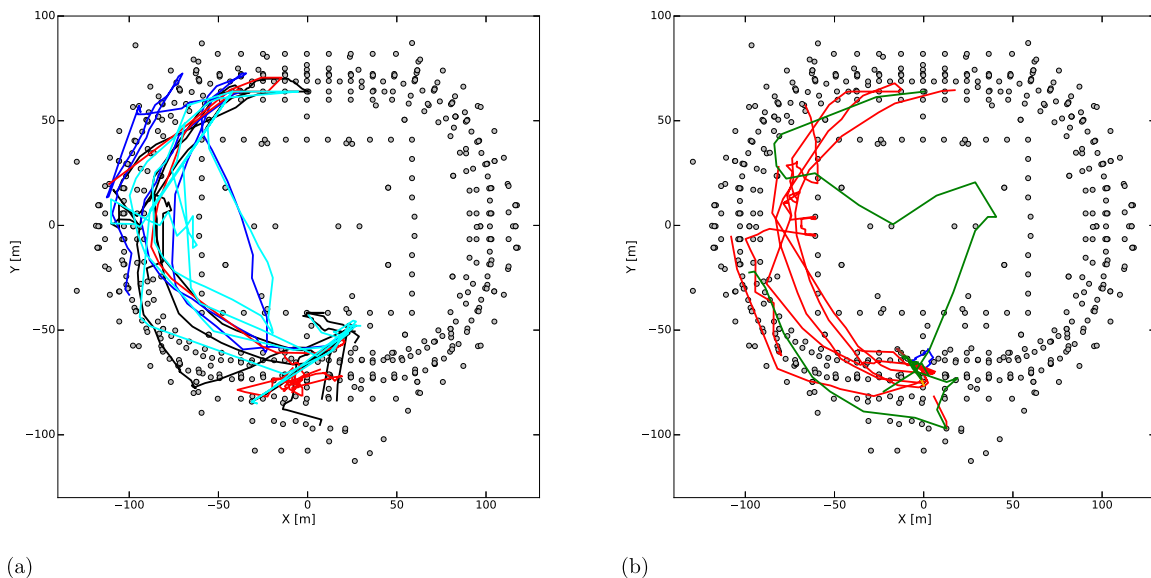
To show how correlated random walks within a bounded area result in truncated superdiffusion, we run a simple simulation. We simulate  $N = 1000$  correlated random walkers for  $T = 200$  time steps. At each time step the individuals make a step of length 10 m, which corresponds to time steps of  $\Delta t = 10$  seconds, and a constant velocity of  $1 \text{ m}\cdot\text{s}^{-1}$ . The correlated random walkers start at random positions and random directions in a rectangular area of  $240 \times 200$  meter, which roughly corresponds to the size of the Arena stadium, and at each step we draw a turning angle from the von Mises distribution with mean  $\mu = 0$  and concentration parameter  $\kappa = 4$  [20]. If an individual's next step ends outside of the rectangular area, the individual remains at the current position and resets its orientation in a new random direction (see Fig. 15 for examples). In Fig. 16 below we see that this simple simulation reproduces some of the characteristics of the empirical MSD shown in Fig. 10 (main text), such as the initial slope of approximately  $\gamma = 7/4$ , and the gradual truncation.

## Appendix D. Statistical independence and group structure

We do not expect that people are going to a dance event alone, but together with friends. Groups of friends moving together during the event will leave similar movement traces, which threatens the statistical independence of the movement data. To test for underlying group structure we apply a cursory form of cluster analysis. The most important aspect of possible group behaviour in our context is the persistent physical proximity of individuals. Therefore, we repeatedly sort individuals based on their pairwise proximity (using a threshold distance  $r = 1.5$  m), creating a sequence of contact networks. We can then search for groups by applying some form of community detection, as is done in various



**Fig. 16.** (a) Empirical MSD on log-log scale (blue line) (same as Fig. 10 and shown here to facilitate comparison). (b) MSD of simulated correlated random walkers (blue line). In both figures the black dashed line corresponds to normal (Brownian) diffusion (proportional to  $t^\gamma$ , where  $\gamma = 1$ ). The solid grey line corresponds to  $\gamma = 7/4$  and is shown as a visual aid.



**Fig. 17.** Movement tracks of core members of potential social groups. The group in (a) has 9 core members, and the group in (b) has 4 members. The movement tracks show similarities but also considerable divergence.

social network studies. Our approach here is similar to the approach presented in Sekara et al. (2016) [37]. If we create contact networks within a sufficiently short time window (called ‘time slice’), individuals are clustered in many small connected components. Thus, within the time slices, communities can be directly observed, which makes more involved community detection algorithms unnecessary.

We define a social group as a community that persists across the time slices. We discard all communities of size  $< 3$  (i.e. pairs of individuals do not count as a group). Communities are matched using single-linkage clustering, a form of agglomerative hierarchical clustering. For the matching, we use a distance measure  $d(c_i, c_j) = 1 - J(c_i, c_j)$ , where  $J$  is the Jaccard similarity between groups  $i$  and  $j$ . The desired result is a collection of clusters consisting of communities that are linked across the time slices. To extract the core members of the group we look for individuals that are present in at least 50% of the lifetime of the community. We assess the results with a graphical check of the group member’s movement tracks.

We experiment with different numbers of time slices and time steps, and find only two potential groups. The results we show here are from taking 20 time slices at a regular time interval of 16 min 40 s (1000 s), starting at 20:06:57, and ending at 01:23:37. We partition the clustering using a threshold distance  $d = 0.5$  and look at communities that are present in at least 3/4 of the time slices. In Fig. 17 we show the movement tracks of the core members of the groups. Note that Fig. 17(a) shows 9 movement tracks, which are in physical proximity mostly when not moving. In both Fig. 17(a)

and (b) movement tracks show similarities but also considerable divergence. Although these results are interesting, we do not find them strong enough to make adjustments to the device selection.

## References

- [1] R. Nathan, W.M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, P.E. Smouse, A movement ecology paradigm for unifying organismal movement research, *Proc. Natl. Acad. Sci. USA* 105 (49) (2008) 19052–19059.
- [2] W.F. Fagan, J.M. Calabrese, The correlated random walk and the rise of movement ecology, *Bull. Ecol. Soc. Am.* 95 (3) (2014) 204–206.
- [3] G.M. Viswanathan, M.G.E. da Luz, E.P. Raposo, H.E. Stanley, *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters*, Cambridge University Press, 2011.
- [4] D.L. Kramer, R.L. McLaughlin, The behavioral ecology of intermittent locomotion, *Amer. Zool.* 41 (2001) 137–153.
- [5] O. Bénichou, C. Loverdo, M. Moreau, R. Voituriez, Two-dimensional intermittent search processes: An alternative to Lévy flight strategies, *Phys. Rev. E* 74 (2006) 020102.
- [6] A.M. Reynolds, On the intermittent behaviour of foraging animals, *Europhys. Lett.* 75 (4) (2006).
- [7] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, *Nature* 439 (2006) 462–465.
- [8] M.C. Gonzalez, C.A. Hidalgo, A.L. Barabasi, Understanding individual human mobility patterns, *Nature* 453 (2008) 779–782.
- [9] I. Rhee, M. Shin, S. Hong, K. Lee, S. Chong, On the Levy-walk nature of human mobility, in: *IEEE INFOCOM 2008 Proceedings*, 2008, pp. 1597–1605.
- [10] D.A. Raichlen, B.M. Wood, A.D. Gordon, A.Z.P. Mabulla, F.W. Marlowe, H. Pontzer, Evidence of levy walk foraging patterns in human hunter-gatherers, *Proc. Natl. Acad. Sci. USA* 111 (2) (2014) 728–733.
- [11] A. Reynolds, E. Cecon, C. Baldauf, T.K. Medeiros, O. Miramontes, Levy foraging patterns of rural humans, *PLoS ONE* 13 (6) (2018).
- [12] L. Alessandretti, P. Sapiezynski, S. Lehmann, A. Baronchelli, Multi-scale spatio-temporal analysis of human mobility, *PLoS ONE* 12 (2) (2017) 1–17.
- [13] X. Liang, X. Zheng, W. Lv, T. Zhu, K. Xu, The scaling of human mobility by taxis is exponential, *Physica A* 391 (2012) 2135–2144.
- [14] G. Riccardi, A. Bazzani, S. Rambaldi, Towards a statistical physics of human mobility, *Internat. J. Modern Phys. C* 23 (9) (2012).
- [15] V. Mendez, D. Campos, F. Bartumeus, *Stochastic Foundations in Movement Ecology: Anomalous Diffusion, Front Propagation and Random Searches*, Springer Complexity, 2014.
- [16] C.T. Brown, L.S. Liebovitch, R. Glendon, Lévy Flights in Dobe Ju/hoansi foraging patterns, *Hum. Ecol.* 35 (2007) 129–138.
- [17] A. Kushki, K.N. Plataniotis, A.N. Venetsanopoulos, *WLAN Positioning Systems: Principles and Applications in Location-Based Services*, Cambridge University Press, 2012.
- [18] D. Douglas, T. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *Cartographica* 10 (2) (1973) 112–122.
- [19] P. Turchin, *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*, Beresta Books, 2015.
- [20] E.A. Codling, M.J. Plank, S. Benhamou, Random walk models in biology, *J. R. Soc. Interface* 5 (2008) 813–834.
- [21] G.M. Viswanathan, V. Afanasyev, S.V. Buldyrev, E.J. Murphy, P.A. Prince, H.E. Stanley, Levy flight search patterns of wandering albatrosses, *Nature* 381 (1996) 413–415.
- [22] A.M. Edwards, R.A. Phillips, N.W. Watkins, M.P. Freeman, E.J. Murphy, V. Afanasyev, S.V. Buldyrev, M.G.E. da Luz, E.P. Raposo, H. Eugene Stanley, G.M. Viswanathan, Revisiting levy flight search patterns of wandering albatrosses, bumblebees and deer, *Nature* 449 (2007) 1044–1048.
- [23] A. Mashanova, T.H. Oliver, V.A.A. Jansen, Evidence for intermittency and a truncated power law from highly resolved aphid movement data, *J. R. Soc. Interface* 7 (2010).
- [24] A. Thiebault, Y. Tremblay, Splitting animal trajectories into fine-scale behaviorally consistent movement units: breaking points relate to external stimuli in a foraging seabird, *Behav. Ecol. Sociobiol.* 67 (2013).
- [25] A. Clauset, C. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, *SIAM Rev.* 51 (2007) 661–703.
- [26] J. Alstott, E. Bullmore, D. Plenz, Powerlaw: A python package for analysis of heavy-tailed distributions, *PLoS ONE* 9 (1) (2014).
- [27] J. Laherrere, D. Sornette, Stretched exponential distributions in nature and economy: ‘fat tails’ with characteristic scales, *Eur. Phys. J. B* 2 (1998) 525–539.
- [28] U. Frisch, D. Sornette, Extreme deviations and applications, *J. Physique I* 7 (9) (1997) 1155–1171.
- [29] R. Gallotti, A. Bazzani, S. Rambaldi, M. Barthelemy, A stochastic model of randomly accelerated walkers for human mobility, *Nature Commun.* 7 (2016) 12600.
- [30] K.P. Burnham, D.R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, second ed., Springer, 2002.
- [31] P.M. Kareiva, N. Shigesada, Analyzing insect movement as a correlated random walk, *Oecologia* 56 (1983) 234–238.
- [32] G.M. Viswanathan, E.P. Raposo, F. Bartumeus, Jordi Catalan, M.G.E. da Luz, Necessary criterion for distinguishing true superdiffusion from correlated random walk processes, *Phys. Rev. E* 72 (2005) 011111.
- [33] J.L. Silverberg, M. Bierbaum, J.P. Sethna, I. Cohen, Collective motion of humans in mosh and circle pits at heavy metal concerts, *Phys. Rev. Lett.* 110 (2013) 228701.
- [34] S. Benhamou, How many animals really do the Levy walk? *Ecology* 88 (8) (2007) 1962–1969.
- [35] A. Reynolds, Liberating levy walk research from the shackles of optimal foraging, *Phys. Life Rev.* 14 (2015) 59–83.
- [36] R. Hilborn, M. Mangel, *The Ecological Detective: Confronting Models with Data*, in: *Monographs in Population Biology*, vol. 28, Princeton University Press, 1997.
- [37] V. Sekara, A. Stopczynski, S. Lehmann, Fundamental structures of dynamic social networks, *Proc. Natl. Acad. Sci. USA* 113 (36) (2016) 9977–9982.