RESEARCH ARTICLE

Monitoring physical distancing for crowd management: Real-time trajectory and group analysis

Caspar A. S. Pouw[1,2], Federico Toschi [1,3], Frank van Schadewijk[2], Alessandro Corbetta [1] *

1 Department of Applied Physics, Eindhoven University of Technology, Eindhoven, The Netherlands, 2 ProRail Stations, Utrecht, The Netherlands, 3 CNR-IAC, Rome, Italy

* a.corbetta@tue.nl

Abstract

Physical distancing, as a measure to contain the spreading of Covid-19, is defining a "new normal". Unless belonging to a family, pedestrians in shared spaces are asked to observe a minimal (country-dependent) pairwise distance. Coherently, managers of public spaces may be tasked with the enforcement or monitoring of this constraint. As privacy-respectful real-time tracking of pedestrian dynamics in public spaces is a growing reality, it is natural to leverage on these tools to analyze the adherence to physical distancing and compare the effectiveness of crowd management measurements. Typical questions are: "in which conditions non-family members infringed social distancing?", "Are there repeated offenders?", and "How are new crowd management measures performing?". Notably, dealing with large crowds, e.g. in train stations, gets rapidly computationally challenging. In this work we have a two-fold aim: first, we propose an efficient and scalable analysis framework to process, off-line or in real-time, pedestrian tracking data via a sparse graph. The framework tackles efficiently all the questions mentioned above, representing pedestrian-pedestrian interactions via vector-weighted graph connections. On this basis, we can disentangle distance offenders and family members in a privacy-compliant way. Second, we present a thorough analysis of mutual distances and exposure-times in a Dutch train platform, comparing pre-Covid and current data via physics observables as Radial Distribution Functions. The versatility and simplicity of this approach, developed to analyze crowd management measures in public transport facilities, enable to tackle issues beyond physical distancing, for instance the privacy-respectful detection of groups and the analysis of their motion patterns.

1 Introduction

Crowd management is a challenging scientific topic directly impacting on the functioning of trafficked urban infrastructures such as, e.g., train or metro stations. Even more so, in time of Covid-19 pandemic, after an initial lock-down period, communities are still wondering how to resume a "new normal" life, while the virus is still circulating among the population. One of

the key control measures has been to maintain a minimal physical distance (often also called "social distance") between any two individuals not belonging to the same family [1]. This distance is country-specific and it ranges from 1m (e.g. China and France), as recommended by WHO, up to 2m (e.g. UK and Canada), being 1.5m in the Netherlands and in some countries it is even adjusted over time. As there is a rather widespread suspicion that we may have to live with such requirements of physical distancing for months to come, it is therefore natural that this is becoming a design requirement for public infrastructures (e.g. [2–5]).

There are however several challenges associated to the automated monitoring of physical distancing in crowds. First, in order to respect individual privacy one needs to employ sensors and techniques that ensure privacy by design while, at the same time, providing accurate space-time information on individual positions with sub-meter accuracy.

Secondarily, one needs to develop algorithms that, while preserving privacy, are capable to autonomously discern, with a good degree of accuracy, families and family members from strangers. This identification should be performed in real-time, raising a number of non-trivial technical challenges.

Additionally, in recent months a number of countries have developed contact tracing apps that allow to receive an alert when somebody has been in "close" contact with somebody that, later on, will turn out to be positive to the Covid-19 [6]. Countries are developing apps based on different alert thresholds, typically a combination of having been closer than a given distance, for a time longer than an established reference. These thresholds, again, are country specific. In Italy and Germany the national apps alert for contacts longer than 15 minutes at a distance below 2 meters. The proper balance of these two aspects, distance and time, is key to avoid too many false positives or false negatives due to e.g. to low or high risk exposures as well as to the inaccuracy of distance estimation via the intensity of Bluetooth signals. It is therefore extremely interesting to be able to analyze, in a number of key urban settings, the combination of contact times for given distances between two persons. This knowledge may provide key information for the calibration of contact tracing apps in different context.

In this paper we employ data from commercial pedestrian tracking sensors placed overhead at Platform 3 in Utrecht central train station (The Netherlands), in order to develop an efficient algorithm, capable of running in real-time, and able to distinguish infringements of the physical distancing rule from the behaviour of family members, that are allowed not to respect such a rule. We introduce the concept of "Corona event", to indicate events when two people, not belonging to the same family, get closer than a threshold distance D.

We focus on contact times and mutual distances considering statistical observables as the radial distribution functions (RDFs), which can conveniently be employed to quantify average exposure times. This enables a two-fold task: automatizing the definition of families and groups (from now on named family-groups) and characterizing the statistical distribution of violations, which we compare with analogous pre-Covid measurements. Based on the space-time dynamics of groups, we try to identify family members as those individuals that consistently stay closer than a given threshold distance for sufficiently long time. This, in turn, allows us to define physical distance violators as those individuals that only occasionally (i.e. inconsistently) yield Corona events infringing the minimal distance rule.

This paper is structured as follows: in Section 2, we survey the pedestrian dynamics literature and computer science methods in connection with group-dynamics and mutual distances. We outline both fundamental outstanding questions and existent analysis methods. In Section 3, we describe the location and measurement setup at Utrecht Central train station used to acquire the analyzed pedestrian data. In Section 4 we review the concept and basic properties of Radial Distribution Functions, extensively leveraged on by our method. In Section 5 we

present our method and possible variations, that we employ to analyze mutual distancing data in Section 6. A final discussion in Section 7 closes the paper.

2 Related works: (Social) distance in pedestrian dynamics

The analysis of pairwise distances and the automated identification of family-groups triggered by the Covid-19 pandemics connect with outstanding technological and fundamental issues in the broader field of *crowd dynamics*. Crowd dynamics is a multidisciplinary research area aiming at understanding and modeling the motion of pedestrians in crowds (see, e.g., [7–9], for introductory references). Outstanding questions specifically connected to mutual distances and groups are, e.g.: "what is the impact of the group on the individual dynamics observables such as position and velocity?" "How do people in social groups interact?", "How does information propagates throughout groups?" (see e.g. [10–14] and e.g. [15] for a group-psychology review). Although these questions are longstanding, and have been investigated via models or in laboratory settings extensively, first quantitative studies in pedestrian dynamics driven by real-life big experimental datasets are relatively recent (see, e.g., [16–19]). Large-volumes of experimental data, in the order of hundred of thousands real-life trajectories, are indeed essential in order to analyze quantitatively and systematically the physics of pedestrian motion, disentangling the high variations in individual behaviors from average patterns, and characterizing typical fluctuations and universal features [19, 20]. This relative delay in performing high-statistics based analyses of pedestrian motion (especially in comparison with other "active matter" physical systems [21]), is most likely due to the complex technical challenge of achieving accurate, privacy-preserving, individual tracking in real-life conditions (see, e.g., [20, 22, 23], or [24] for approaches targeting even higher resolution). Market solutions, as the one considered in this paper, are also becoming accessible, offering various trade-offs between accuracy and costs (see, e.g., [25]).

On top of automated tracking, higher-level automated understanding of individual behaviors—a concept also known in computer science as trajectory pattern mining [26, 27]—remains also outstanding in many aspects. The automated identification of pedestrian groups, or pedestrian "group mining", is a notable example in this context. On one side, in current pedestrian dynamics research, the definition and classification of groups and social structures in experimental data has been manual, i.e. based on labor-intensive visual inspection (e.g. [28]). While this ensures high-quality validated measurements, it limits the possibility to establish vast statistical datasets towards data-driven characterizations of averages and fluctuations in the dynamics. On the other side, automatic strategies to identify groups have been proposed by the data mining community. These approaches primarily hinge on analyzing (instantaneous) spatial clusters of pedestrians and the consistency with which these adhere over time to some group semantics (flocking, convoying, aggregation/ desegregation, see [27, 29–31] and references within).

Here we pursue distance analyses and family-group identification via discretized mutual pairwise distance distributions—represented in physics terms via Radial Distribution Functions among relevant pedestrian pairs (the concept of RDF is further reviewed in Section 4). We accumulate information on a "social" interaction graph with vector edge weights. This data structure holds all the relevant contact times and distance statistics; besides, family-groups emerge as incremental features queryable by a space-time distance semantics. Graphs are classic tools in discrete mathematics to represent networks of interactions, or connections between entities (e.g. [32]). Formally speaking, a graph H is a set of nodes, $H = \{p_i\}$, endowed with edges, say $e = (p_i, p_j)$, connecting node pairs. Providing a weight function, $w(e)$, defined on the edges, makes the graph "weighted". In our case, nodes are in 1:1 correspondence with observed

pedestrians, whereas edges underlie distance-based interactions, that are characterized by a weight function with values in a real vector space of pre-fixed dimension. Graphs have been often used for data-driven studies on social behavior both of humans, e.g. to analyze social networks [33], GPS-data [34], but also of social animals (e.g. [35, 36]). In [19, 37], graphs have also been used to address big-data analyses and representation of pedestrian dynamics aiming at efficient data searches.

3 Pedestrian tracking setup at Utrecht Central station

We benchmark our approach considering pedestrian tracking data acquired on platform 3 at Utrecht Central station, The Netherlands (cf. Fig 1). Utrecht Central, with roughly 57 million annual users, is the nation-wide busiest railway station. Since 2017, platform 3 has been equipped with 19 commercial pedestrian tracking sensors, each of which captures 3D stereo images at $f = 10$ frames per second and processes them to deliver individual tracking data in a privacy-friendly manner (cf. sketch in Fig 1). The sensor view-cones are in partial overlap, which enables the sensor network to stitch together trajectory pieces acquired by the single devices. The total area covered by the set of sensors consist of the full platform width (about 3m) for 120 linear meters next to track 5, plus the area underneath escalators and staircases connecting the platform to the central hall. This yields a covered area of approximately 450m^2. Track 5 is among Utrecht's busiest tracks and is primarily utilized by trains heading to Amsterdam Central Station and Schiphol Airport. The complex and multi-directional crowd flows on the platform are recorded with high space- and time-resolution 24/7 since March 2017 (localization precision: $O(5\text{–}10)$cm, similar technology to what employed in [25]). In normal operation conditions, the system would capture about 100.000 trajectories per day while, on average, only 16.000 trajectories per day were observed in the two months after the Covid-19 outbreak (see pre- and during- Covid-19 crowd distribution example in Fig 1 and crowd density histograms in Fig 2). This unique measurement setup gives us not only the possibility of developing and testing our approach in meaningful real-life conditions, but also to compare relevant statistical observables of the pairwise distance (RDFs), before and during the Covid-19 measures.
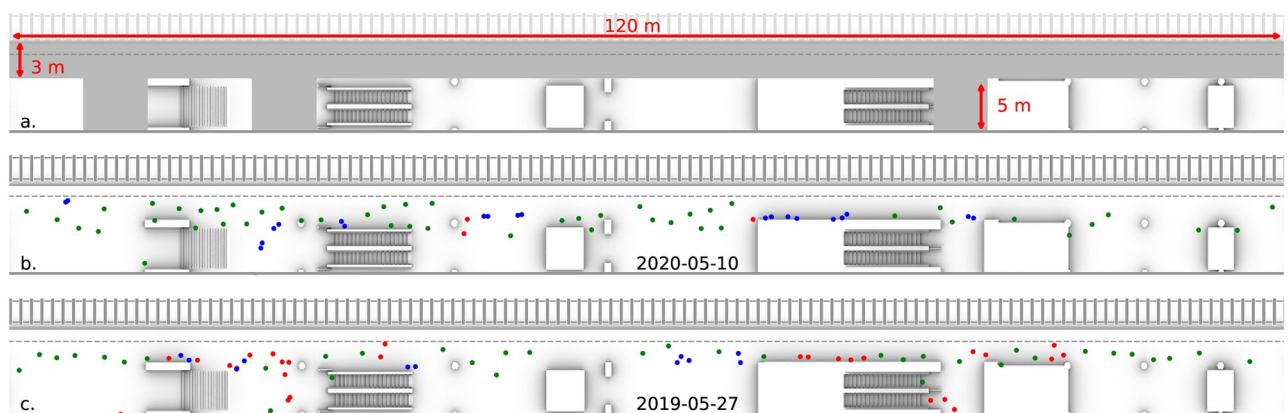
Fig 1. (a) Floorplan of platform 3 at Utrecht Central Station (NL). The area monitored by the sensors is highlighted in grey. (b) Sample of 75 passengers waiting for a train to arrive on the 10th of May 2020, during the Covid-19 pandemic. Pedestrians which respect the 1.5m physical distance regulations are colored in green, people who are part of a family-group are colored in blue, distance offenders are colored in red. This classification is performed via the method proposed in Section 5. In this situation only 3 out of the 75 people violate the physical distancing rules. (c) Same number of people distributed over the platform on the 27th of May 2019, one year prior to the Covid-19 outbreak, here about one-third of the people stand closer than 1.5m to someone else.

https://doi.org/10.1371/journal.pone.0240963.g001
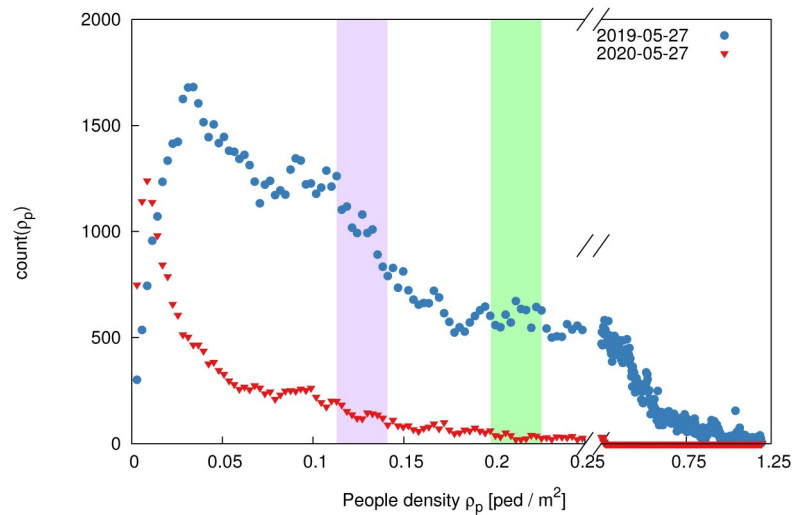
Fig 2. Histogram of observed crowd density levels comparing a day before the Covid-19 outbreak (27th of May 2019, blue dots) and for a day during the Covid-19 pandemic (27th of May 2020, red triangles). Prior to the Covid-19 outbreak, densities in excess of 1ped/m² occurred daily. One year later, during the Covid-19 pandemic, the maximum crowd density recorded is only about 0.3ped/m². We compare measurements acquired at similar density levels, i.e. where the average available space per person is comparable. We focus on two density levels: 40-50 passengers (purple band, cf. Figs 3(a), 3(c) and 6(a)) and 70-80 passengers (green band, cf. Figs 1, 3(b), 3(d) and 6(b)).

Ethics statement

This study has been approved by the Ethical Review Board of Eindhoven University of Technology (ref. ERB2020AP1, 21st, Feb. 2020). During this study we acquired passengers trajectory data (Utrecht station, Platform 3) via anonymous-by-design commercial sensors. The dataset considered includes thus individual trajectories and no additional personal feature has been available and/or stored.

4 Pedestrian radial distribution functions

In theoretical physics and molecular dynamics, the radial distribution function, $g(r)$ (RDF), and the radial cumulative distribution function (RCDF), $G(r)$, are established tools to characterize the distribution of pairwise distances between particles (see e.g. [38]), i.e., in our case, pedestrians.

By definition of RCDF, for a crowd with uniform spatial density ρ, on average, i.e. in the mean-field of many realizations, the number of people, $N_\rho(r)$, at a distance *up to r* from a generic individual satisfies

$$N_\rho(r) = \rho G(r) = \rho \int_0^r g(z)\, dz, \tag{1}$$

therefore, $g(r) = \partial_r G(r)$ holds. Thus, the functions $g(r)$, $G(r)$ (and derived quantities) do not carry any space/time specific information, rather they relate to average properties, depending only on mutual distances.

For instance, in unconfined space, $G(r)$ grows as the circle area, i.e.

$$N_\rho^{unconf}(r) = \pi r^2 \rho. \tag{2}$$

In our train platform, such a $\sim r^2$ growth ratio is possible only when r is sufficiently smaller
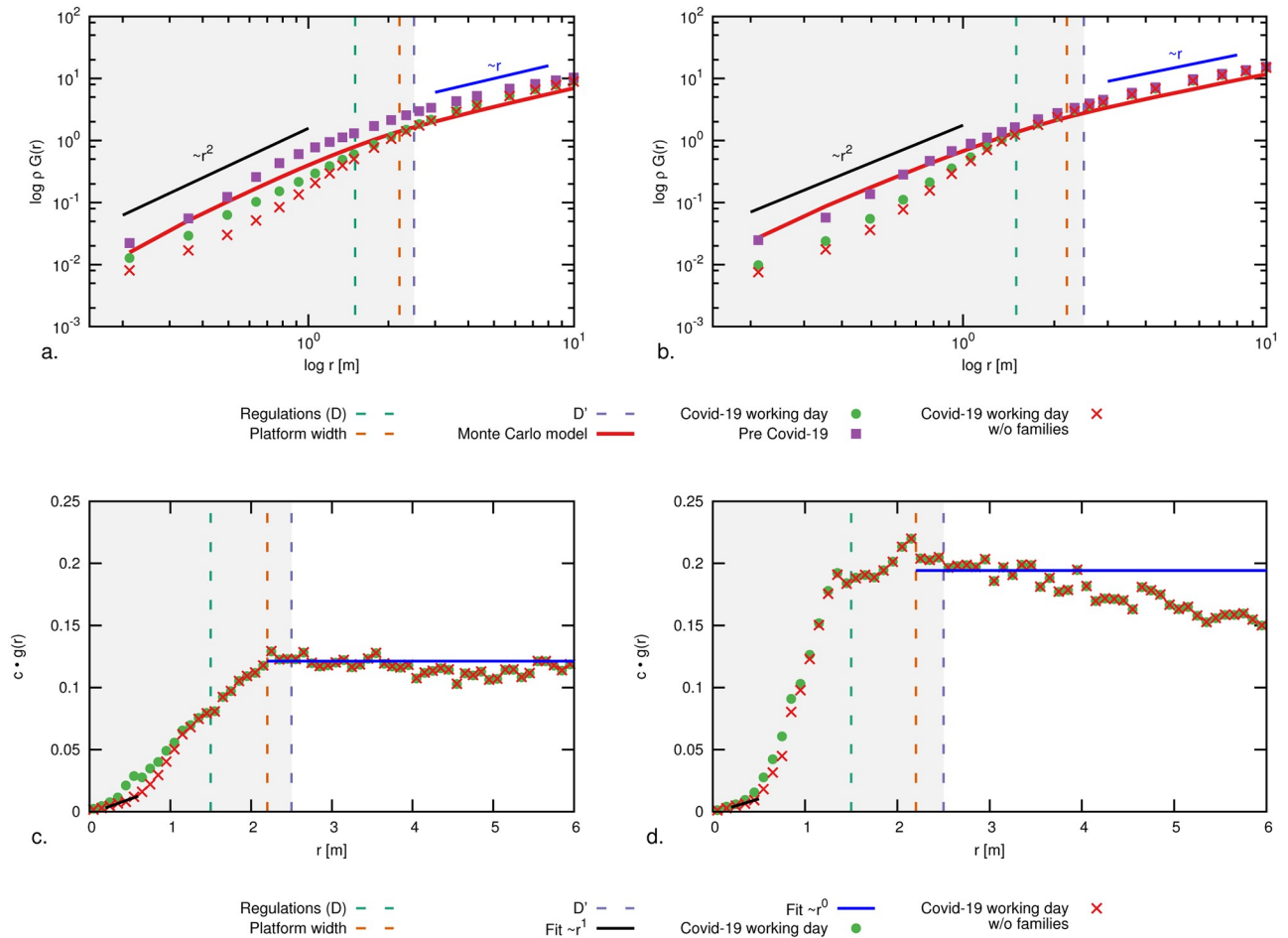
Fig 3. (a, b) Radial cumulative distribution functions (RCDF), $g(r)$, and (c, d) radial distribution functions (RDF), $G(r)$, for density levels on a typical working day. On the left (a, c) for density level 1, with 40-50 pedestrians on the platform, (green domain in Fig 2) and on the right (b, d) for density level 2, with 70-80 pedestrians on the platform (purple domain in Fig 2). Vertical dashed lines at 1.5m, 2.2m and 2.5m indicate, respectively, the Dutch social distancing regulations ($r < D$), the usable width of the platform (without danger zone) and the critical threshold D'. The solid black line at small r values highlights the $\sim r^2$ growth ratio up to 2.2 m and a blue line for the $\sim r^1$ trend at larger mutual distances. In (b,d) the normalization constant c is chosen such that $\int_0^\infty cg(z)\,dz = N$, where N is the number of people on the platform. Similar plots for a weekend day are reported in Fig 6. We compare the pre-Covid situation with the present, and with a Monte Carlo model of a random distribution of passengers across a region identical to the platform. We report the RDF and RCDF of the current situation including and excluding family-groups contributions, as made possible by the method introduced in Section 5.

than the platform width. Else, we expect a $\sim r^1$ (linear) trend, as the area growth is bound to the platform length only. This holds until platform finite size effects come into play. In Fig 3 we compare the RDF and RCDF for the two density levels highlighted in Fig 2. We notice a depletion in the radial distribution functions at short distances when comparing with the situation pre-outbreak. As a partial anticipation of the results of this paper, in the figure we report also the RDF discounted of short-distance family-group interactions (such interactions are allowed by present regulations). As expected, this yields a further depletion of the RDF in the region $r \lesssim 1.5$ m. In the figure, we additionally report the aforementioned analytic trends and the R(C)DF functions obtained through Monte Carlo numerical simulations in a rectangular domain with the same size as the platform. We specifically consider an ensemble of simulated crowds of N pedestrians; individuals have a random spatial distribution satisfying a minimal mutual distance of 0.2m. The figure reports ensemble-averaged RDFs and RCDFs which thus

include small-range quadratic growth, platform-width-bound linear growth and finite size effect. Measurements well-conform with simulations.

Short mutual distances over extended time duration are known to increase the contagion probability: RDF and RCDF can be used to evaluate the average exposure time. A pedestrian that was on the platform for a time interval ΔT, was exposed, on average, for a time

$$T^{r_c} = \Delta T \rho G(r_c), \tag{3}$$

with r_c being a critical distance threshold (e.g. r_c can be the nation-wide physical distancing requirement). Similarly, the function

$$t(r) = \Delta T \rho g(r), \tag{4}$$

quantifies the contribution to the total average exposure time given by peer pedestrians at distance r.

5 Distance-interaction network

In this section, we describe our scalable framework to characterize pedestrians pairwise distances, identify family-groups and distance offenders. Our approach is a substantial evolution of the graph approach presented by some of the authors in [19, 37], which not only does enable to tackle, in real-time, relevant questions connected to Covid-19 safety measures, but unlocks the automation of relevant pedestrian dynamics aspects as the detection of groups. Relevant differences with the previous works [19, 37] are postponed after the approach description.

Our measurements come in the form of time-stamped trajectories. As no further information is available, such as body orientation or gaze direction [39] or body size/approximate age, our identification of family-groups relies only on mutual proximity and its time-consistency. Whenever two or more pedestrians maintain a mutual short distance consistently throughout a sufficient fraction of their trajectory, they should automatically emerge as belonging to the same family-group. Additionally, we deem implementation simplicity, while maintaining efficiency and sufficient accuracy in identifying family-groups relations, possibly in real-time, and without complex/costly data-searches. Hence, our approach is "additive" (or "incremental") and RDF-like information is increased, on the go, in a graph data structure (at minimal memory costs), and, without computationally-costly searches in stored records, family-groups and offenders remain identified immediately. In other words, by additivity, we stress that our data structure is built online and usable after only one time-forward pass of the trajectory data.

5.1 Graph data structure construction

In conceptual terms, we represent the pedestrian trajectories as individual nodes of a graph H. Each node includes information specific to the trajectories, such as overall observation time, τ, source and destination. These three quantities are incremental as, respectively, τ scales with the number of frames a pedestrian is observed, the source point is the initial point of a trajectory while the destination gets constantly updated with the current position until a pedestrian leaves the measurement area. Whenever two pedestrians, say p_1 and p_2, are observed simultaneously (i.e. in the same frame) and their Euclidean distance, $r = d(p_1, p_2)$, is below a critical threshold $D' > D$, we memorize (properties of) this event within the weight, $\vec{w}(e)$, of the edge $e = (p_1, p_2)$, that connects the two pedestrian-nodes p_1, p_2. Specifically, the weight $\vec{w}(e)$ aims at a discrete counterpart of the RDF ($g(r)$, cf. Eq (1)) restricted to pedestrians p_1, p_2 and with support $0 \le r \le D'$. Similarly to the RDF, also the graph H does not hold detailed microscopic information in space/time, such as instantaneous positions. To focus on the identification of Corona events and disentanglement of family-groups, we set $D' = 2.5$ m, i.e. one meter more

than the social distance required by Dutch regulations, $D = 1.5$ m$< D'$. This aims at exploring the distance dynamics in the neighborhood of the current regulations and leaves flexibility should the regulations become stricter and require additional mutual separation.

The vector weight $\vec{w}(e)$ keeps record of the number of occurrences of distance events, r, after a given radial quantization (binning). In the following, we consider five evenly-sized bins with sides at

$$r = \{r_0, \ldots, r_d, \ldots\} = \{0, 0.5, 1.0, 1.5, 2.0, 2.5\} \text{ m}; \tag{5}$$

for the sake of readability, we also indicate, the individual components of $\vec{w}(e)$, $\vec{w}_d(e)$, as

$$\vec{w}_0(e) = \vec{w}(e)[0.0 : 0.5],$$
$$\vec{w}_1(e) = \vec{w}(e)[0.5 : 1.0],$$

and so on. Hence, for each time instant in which $d(p_1, p_2) < 0.5$, the counter $\vec{w}(e)[0 : 0.5]$ is incremented of one unit. Similarly, whenever $0.5 < d(p_1, p_2) < 1.0$, the weight $\vec{w}(e)[0.5 : 1.0]$ gets incremented, *etc*. Note that updating the data structure requires only all the pairwise distances (smaller than D') at each frame. The choice of bin size, which regulates the quality of the approximation of the RDF of p_1 and p_2, is clearly arbitrary and needs to be a trade-off between the required resolution on the RDF and memory allowance.

Consistently with Eq (4), scaling the counts $\vec{w}(e)$ by the (inverse of the) sensors sampling frequency, f, yields the time duration, in seconds, pedestrians p_1 and p_2 maintain a given (quantized) distance (i.e. $f^{-1}\vec{w}(e)[0.5 : 1.0]$ is the amount of seconds p_1 and p_2 had a distance between 0.5 and 1.0 meters). Hence, statistical moments of r, weighted by $f^{-1}\vec{w}(e)$ enable to calculate the total contact time of p_1 and p_2, their average distance and fluctuations. In all cases, the statistics are restricted to $r \leq D'$ (insights on the relevant statistical properties of the graph are left to the next subsection). Operationally, we build the graph as reported in Algorithm 1. Additionally, in Fig 4a, we provide a visual description of the graph in the case of a subsection of our train platform, while in Fig 4b we show examples of typical graphs built in time windows about 10 minute-long around the train arrivals.
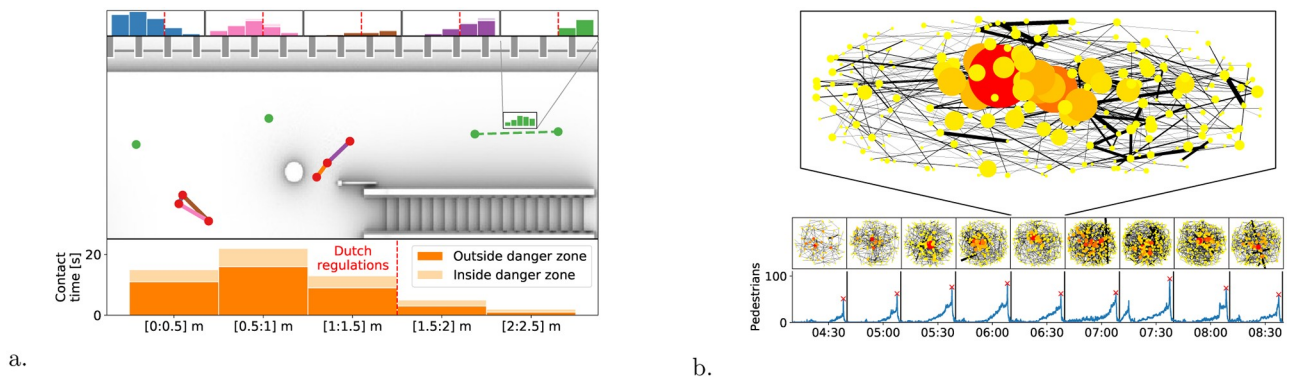
Fig 4. (a) Conceptual sketch representing the accumulation of information on the graph H. Whenever two pedestrians, say p_1, p_2 stand at a distance d smaller than D', this gets recorded in the histogram weight of the edge between nodes p_1 and p_2 as an additive contribution to the bin approximating d. In the sketch we report a section of the platform: edge appear between nodes according to the distance; the histogram weights are reported atop and beneath the sketch with the same color coding of the edges and scaled with the sampling time (thus they translate to the contact time conditioned by the distance). Nodes are reported in red if they have performed at least one Corona event (thus they have an edge with non-zero contributions at distances below D'), else they are in green. (b) Examples of graphs acquired in windows of about ten minutes around each train arrival (determining the peaks in the counts at the bottom). We report a magnified version of one among these graphs. Nodes are colored by the node degree, i.e. by the number of first neighbors, ranging from yellow to red. Edge thickness scaled by the contact time, T_e^d, Eq (7). The higher the degree of a node, the larger the number of distance offenses performed by the associated pedestrian.

Algorithm 1: Graph construction algorithm in pseudo-code. The data structure is built streaming the trajectory data once. The function $q = q(d)$ returns the distance bin to which d belongs. Hence, given the quantization in Eq (5), it holds $q(0 < d < 0.5m) = 0$, $q(0.5m < d < 1.0m) = 1$, etc.

```
Data: Trajectories dataset, possibly live-streaming
Result: Distance-interaction graph H
H = empty graph;
for t in time do
  add any trajectory, pᵢ, starting at time t as a node in H, store
origin;
  update persistence time τₚᵢ, destination of all observed trajectory-
nodes pᵢ;
  for 0 < i < j ≤ # trajectories at time t do
    if d(pᵢ, pⱼ) < D′ then
      w⃗(pᵢ,pⱼ)[q(d(pᵢ,pⱼ))]+ = 1
    end
  end
end
```

Extensions and variations.   In settings as train platforms, not all the areas come with the same importance or criticality. The so-called "danger zone", the last 80cm-wide buffer region on the platform that is stepped on just before boarding a train, is an example. For our use case it is imperative to have the capability of discriminating between events happening inside and outside such an area. To achieve this, we consider two separate sets of weights on each edge: $\vec{w}^{dz}(e)$ and $\vec{w}^{c}(e)$, which count, respectively, the time instants a given distance below D' occurs when the centroid between the two pedestrians lies in the danger zone, and otherwise. According to our previous definition, $\vec{w}(e) = \vec{w}^{dz}(e) + \vec{w}^{c}(e)$ holds. Fig 4a reflects this aspect by representing $\vec{w}^{dz}(e)$ and $\vec{w}^{c}(e)$ stacked (and with different color shade).

5.2 Approximation of the short-range RDF as edge average

The graph is a collection of RDFs functions restricted to pairs of pedestrians. As such, within the limits of the quantization, it is richer in information than the "global" RDF (Eq (1)). The latter, in fact, can be recovered by averaging the edge weights over the entire graph, i.e. by combining each pairwise contribution. Restricting to a graph describing conditions with equal density, the global RDF can be approximated as

$$\int_{r_d}^{r_{d+1}} g(r)\, dr \approx c \left\langle \frac{\vec{w}_d(e)}{\sum_j \vec{w}_j(e)} \right\rangle_e , \quad \forall d, \tag{6}$$

where c is a constant scaling depending on the normalization considered for g. In words, the integral of the RDF in the bin $[r_d, r_{d+1}]$ can be approximated by the d-th (enseble-averaged) edge weight. Averaging over a graph including non-homogeneous density levels, yields the RDF averaged among such densities.

5.3 Interaction classification

In this subsection we leverage on the graph topology and edge data to deduce relevant information about pairwise distance, family-group relations, exposure times, and physical distance offenders.

Pairwise exposure time and pairwise distance statistics.   Pedestrians p_1 and p_2, whose distance satisfied $r = d(p_1, p_2) \leq D'$ at least in one frame, have their interaction recorded on the graph edge $e = (p_1, p_2)$. The weight $\vec{w}(e)$ allows us to characterize their distance properties

restricted to the instants in which $r < D'$. In particular, the *contact time* T_e between p_1 and p_2 satisfies

$$T_e^d = f^{-1} \sum_{j=0}^{d} \vec{w}_j(e), \qquad [\text{Contact time}] \tag{7}$$

where the index d selects the farthest relevant distance bin. According to our quantization in Eq (5), $d = 2$ would restrict to interactions with $r \leq 1.5$ m and thus quantify the "exposure time" according to the Dutch regulations, whereas $d = d_{max} = 4$ includes all stored interactions for the pair e, i.e. $r \leq D'$. Similarly, the average pairwise distance reads

$$\langle r \rangle_e^d = \frac{f^{-1} \sum_{j=0}^{d} r_{j+\frac{1}{2}} \vec{w}_j(e)}{T_e^d}, \qquad [\text{Avg. pairwise dist.}] \tag{8}$$

where $r_{j+\frac{1}{2}}$ identifies the mid point between bin j and $j + 1$ in Eq (5). Higher order moments of r weighted by $\vec{w}(e)$ can be used to estimate the fluctuation (variance) of the distance in time. In particular, if the variance $\sigma^2(r) = \langle r^2 \rangle_e^d - (\langle r \rangle_e^d)^2$ is small, then the pedestrians kept an almost fixed distance during their trajectories. Notably, small or possibly zero r variance, $\sigma^2(r)$, are necessary condition for non-positive (Finite Time) Lyapunov Exponent of the distance between p_1 and p_2 [40].

Total individual exposure time. The total time an individual p has been exposed to contacts can be computed by summing the pairwise contact times T_e^d (Eq (7)) for all pedestrians, p_j, that entered into contact with p, i.e. for all the edges $e = (p, p_j)$ that converge to p, in formulas

$$T_p^d = \sum_{p_j \in N(p)} T_{(p,p_j)}^d, \qquad [\text{Individ. expos. time}] \tag{9}$$

where $N(p)$ is the list of the first-neighbor of p (nodes connected to p through at least a single edge). Eq (9) provides a counterpart to Eq (3) in which we consider a specific pedestrian, p, rather than averaging over all pedestrians. Notice that the index d is the discrete analogue of the cutting threshold r_c.

Family-group relations. We determine whether two pedestrians, p_1, p_2, belong to the same family-group on the basis of their contact time $T_{(p_1,p_2)}^d$ and their persistence time in the tracking area, τ_{p_1} and τ_{p_2}. In particular, if the symmetric relation henceforth indicated as $p_1 \sim p_2$ holds

$$p_1 \sim p_2 \Leftrightarrow \qquad [\text{Fam} - \text{group condition}]$$
$$\min\left(\frac{T_e^{(1)}}{\tau_{p_1}}, \frac{T_e^{(1)}}{\tau_{p_2}}\right) > \lambda^{(1)} \quad \text{and} \quad \min\left(\frac{T_e^{(2)}}{\tau_{p_1}}, \frac{T_e^{(2)}}{\tau_{p_2}}\right) > \lambda^{(2)}, \tag{10}$$

we consider p_1 and p_2 as belonging to the same family-group. We set $\lambda^{(1)} = 40\%$, $\lambda^{(2)} = 90\%$ which, in words, translates to people who have a pairwise distance of less than 1.5m for 90% percent of the time and are within 1m for 40% percent of the time. The rationale being that pedestrians who followed the same trajectory, thereby being in mutual close proximity for the major part of their persistence time, and who are comfortable for extended periods in each other's private space ($r \leq 1$m) most likely belong to the same family-group. Family-groups with more than two individuals are expected to appear in the graph as completely connected sub-graphs, or cliques, in which all the nodes are in Relation (10) between each other.

We can now define a second total individual exposure time which is equal to T_p^d (Eq (11)) after discounting family-group contacts:

$$T_{p,f}^d = \sum_{\substack{p_j \in N(p) \\ p_j \nsim p}} T_{(p,p_j)}^d. \qquad \text{[Individ. expos. time w/o families]} \tag{11}$$

Analogously, we can consider a RCDF discounted of family-group contributions, say $G_f(r)$ (cf. Eqs (1) and (3)), such that

$$T_f^{r_c} = \Delta T \rho G_f(r_c) \tag{12}$$

is the total exposure time with non-family individuals and up to a spatial threshold r_c. By differentiation (as in Eq (1)), we can similarly define the RDF $g_f(r)$ discounted by family-groups.

Family sub-graph transitive closure. The relation "$\sim$" can be non-transitive, i.e. $p_1 \sim p_2$, $p_2 \sim p_3$ do not imply $p_1 \sim p_3$. One may thus consider the transitive closure of "$\sim$", say "" which is defined as p_1 and p_3 belong to the same family-group ($p_1\bar{} p_3$) either if $p_1 \sim p_3$ or if they have common family-group members.

In the data we analyze in Section 6, we do not consider such transitive closure. In other words, family-group relations are exclusively defined by Relation (10). On one hand, we deem rare the event that a family remains not represented by a clique. Even in this case, we expect contributions to the overall RDF statistics be minimal. On the other hand, in real-life data collected with sensors similar to ours, short/broken trajectories may appear. We observed that these, in combination with Relation (10) would unrealistically increase the probability of observing large family-groups. Hence, to avoid excessive and unjustified complications in the heuristics we restrict to Relation (10).

Relevant interactions, family-discounted graph and offenders. Combining the previous elements we can now identify distance offenders as those pedestrians that violate physical distancing while not being part of a family. We consider the sub-graph $H' \subset H$ obtained after pruning H of family-group edges. The connections left in H' must indicate sporadic (i.e. not time-consistent) distance infringements.

As exposure time is deemed a critical parameter for contagion [41], we apply a further time requirement to discriminate actual offenders. Specifically, we introduce the set P'_α defined as

$$P'_\alpha = \{p \in H' : T_{p,f}^d > \alpha\}; \tag{13}$$

in words, elements of P'_α are pedestrians who violated physical distancing with non-family members for an overall time longer than α. The number of first neighbors of a node in P'_α identifies how many contacts such pedestrian had: we label as repeated offenders those with more than 10 first neighbors (i.e. pedestrians that violated physical distancing with more than 10 different people and for an overall time larger than α). We remark that this classification can be run in real-time as all the aforementioned requirements can be constructed in additive manner.

5.4 Differences with previous works by some of the authors

In [19, 37] some of the authors leveraged on a similar interaction network to build a scalable representation and search tool for high-statistics real-life pedestrian tracking data. The aim there was to tackle fundamental issues about the physics of pedestrian dynamics (e.g., mechanics of pairwise avoidance, statistical observables of undisturbed pedestrian motion), which requires extracting pertinent data from databases with few million real-life trajectories. This was made possible and efficient by representing every target experimental scenario with a sub-

graph topology. Matches of this topology within the experimental data correspond to the desired trajectory data. The graph structure hereby proposed focuses, instead, on processing streaming data for RDF and exposure time statistics. As a by-product, we automate detection of groups, in our case mostly families, and related properties.

6 Physical distancing at Utrecht Central station, platform 3

In this section we employ the graph H to analyze trajectory data collected in Utrecht station (see Section 3) and we compare statistics from before and during the Covid-19 pandemic.

Typical graphs and qualitative aspects

In Fig 4b, we report examples of the graphs acquired during a typical morning (time interval 4AM—8.30AM). Train arrivals are the most critical conditions when it comes to respecting physical distancing, thus we create a new graph two minutes after each train departure, when the platform is almost empty (this step is not strictly necessary, but increases computational efficiency). In the figure, nodes size and color follows the node degree, i.e. the number of first neighbors and, thus, the distance offenses committed by that node.

As a qualitative example of the capability of the method to extract relevant data, we show-case two antipodal conditions in Fig 5. In the first case (left panel), we report two pedestrians in a family-group relation that remain together throughout their entire trajectories: from the escalators to the boarding. In the second case, we have a repeated offender: the associated node exhibits 28 first-neighbor connections. Interestingly, a significant part of the offenses happens while the pedestrian waits in proximity of the escalator. This, therefore, rather marks a waiting area to be disallowed, than a willing offender.

Family-group discounted RDFs

In Figs 3 and 6 we report RDFs prior and after excluding family-group interactions. The RDFs for $r < D$ are non-vanishing, even after discounting the contributions of family-groups, most significant in the weekends (when the presence of workers and commuters is lower; cf. short-distance "bump" in Fig 6). We expect these remaining contributions at $r < D$ to be to Corona events by distance offenders. Notably, once removed of family-groups contributions, the RDFs at small r values recover a linear growth rate, as expected by a random spatial distribution of passengers (scaling as the derivative of Eq (2)).

a.

b.
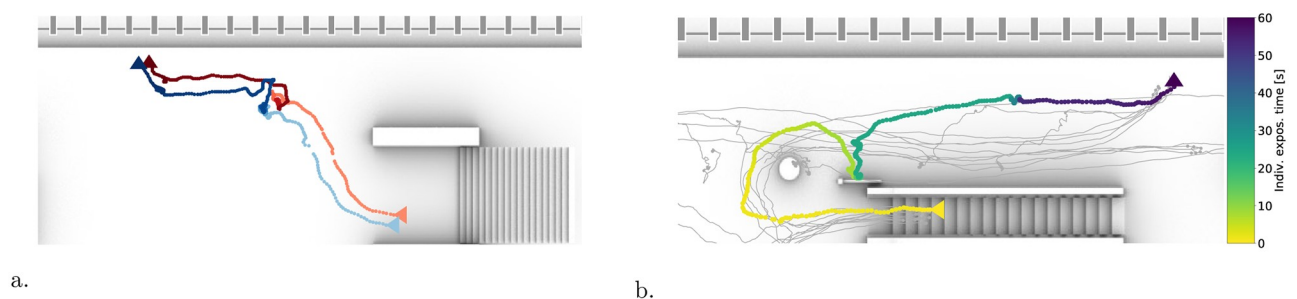
Fig 5. (a) Detected clique consisting of two nodes representing two people traveling together. Both entering the platform through the stairs, waiting together for the next train to arrive and finally boarding the train through the same door. The hue of the trajectories is proportional to the time spent on the platform. Lighter hue when the people enter and a darker hue when they leave. Jump in hue indicating the place where the travelers were waiting. (b) Detected node with degree higher than 10, i.e. a repeated offender who violates physical distancing with more than 10 other people. The trajectory of the repeated offender is reported in shades scaled to the exposure time, while the trajectories of other people that were met violating physical distancing are in gray. The considered offender entered the platform via the escalators and waited underneath the escalators for their train.

https://doi.org/10.1371/journal.pone.0240963.g005
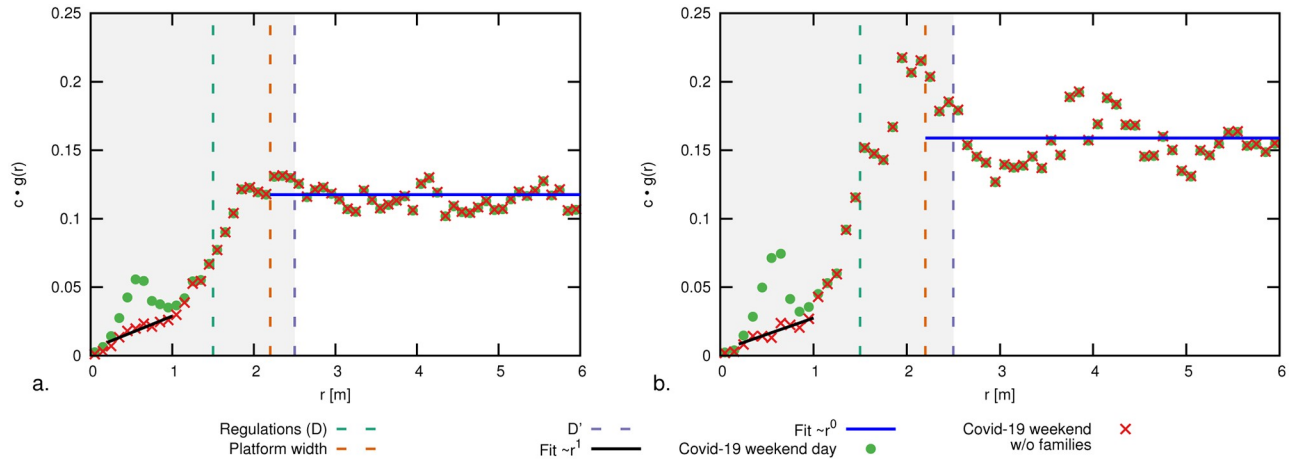
Fig 6. Radial distribution functions (RDF), $g(r)$, for a typical weekend day in case of (a) 40-50 pedestrians on the platform, (green domain in Fig 2) and (b) 70-80 pedestrians on the platform (purple domain in Fig 2). The same conventions of Fig 2 hold. The presence of family-groups determine a peak in the RDFs around $r \approx 0.5$m, which is much more pronounced than in the working day case. Discounting these contributions via the graph analysis notably restores a $\sim r^1$ growth rate at small r values.

https://doi.org/10.1371/journal.pone.0240963.g006

Exposure time, node degree and evolution through the pandemics

In Fig 7a we report, week-by-week, the average edge weight as introduced in Eq (6), pruned of family-group contributions and scaled by the sampling frequency. This provides an approximation of how the (family-discounted) average individual exposure time has been changing over time (weeks 17 to 26 in 2020). As the usage of the platform grew after a drop at the beginning of the outbreak, so did the exposure time for distances between 0.5m and 2.5m, especially until week 22. On the opposite, the amount of time, per person, spent with a peer within 0.5m has remained constant and within fractions of a second. A new operation schedule at week 23,
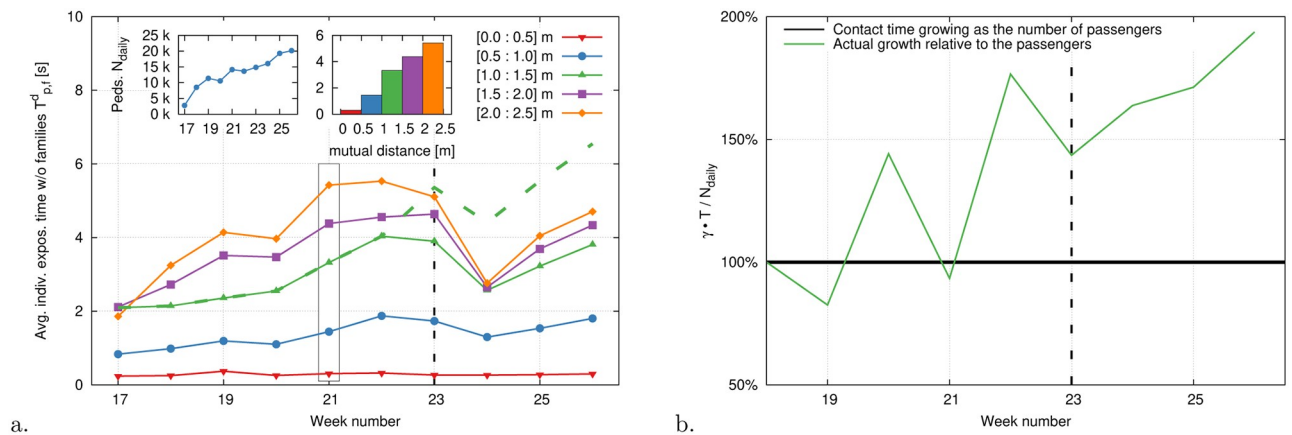
Fig 7. (a) Average individual exposure time without family contributions (Eq (9); weeks 17-26, working days only. Corona lockdown measures in The Netherlands started around week 13). Each line reports average data from bins $\vec{w}_0(e)$, $\vec{w}_1(e)$, etc. The inset on the left shows the average daily passenger count, N_{daily}. The inset on the right reports the same individual exposure time data for week 21 in histogram form. A change in the train schedule on the 2nd of June (week 23, indicated with a vertical black dashed line) increased the train frequency by a factor $\gamma \approx 1.7$ (Eq 14). This improved the distribution of pedestrians over the day thereby temporarily decreasing the individual exposure time. To make the data comparable over time and compensate for the train increment, we multiply the exposure times by γ, Eq (14), (dashed green line, bin $\vec{w}_2(e)$ only, i.e. $r \in [1.0, 1.5]$ $m \approx D$). We notice that the compensated exposure time grows steadily in time gaining a factor 3.5×. This is a combined effect of the passenger growth and a reduction in attention and/or difficulty in adhering to physical distancing regulations. In panel (b) we scale the exposure time by the number of passengers, i.e. we report $\gamma T_{p,f}^d / N$ ($d = 2$). This ratio, which we further scale to its value at week 18, displays a $\approx 100\%$ growth between week 18 and 26, to confirm that the increment of passengers contributes only for 150% of the overall exposure time growth.

https://doi.org/10.1371/journal.pone.0240963.g007

with increased train frequency throughout the day, allowed a temporary reduction of the load on the platform, making easier to respect physical distancing. From week 24 onward, individual exposure times showed again a growing trend. To render the data comparable, we consider also exposure times compensated for the new train schedule (Fig 7a, only for the case $r \approx D$), i.e. corrected by a factor

$$\gamma = \frac{\text{\# daily trains with new schedule}}{\text{\# daily trains with old schedule}} \approx 1.7. \tag{14}$$

This shows a more stable growth pattern and an increase of factor 3.5 between weeks 17 and 26 (the factor γ is an estimate, considering the presence of trains of different kind and lengths). Scaling the corrected exposure times with the number of passengers, which is itself growing, we additionally notice, that the former is growing faster (i.e. exposure time grow super-linearly with respect to the passengers). This suggests a possible relaxation or an increased difficulty in following physical distancing regulations.

We report in Fig 8a an in-depth breakdown of the distribution of node degrees, i.e. the number of first neighbors of each node and thus the number of contacts with different individuals the node had (including both offenses and families). Consistently with our previous observations, the fraction of high degree nodes (5+ or 10+), i.e. repeated offenders, has also been growing steadily, but a temporary drop following the train schedule change. In Fig 8b we display the distribution of individual exposure times pruned of family contributions, and up to the distance thresholds r_d (Eq (5)), i.e. the pdfs of $T_{p,f}^d$ (cf. Eq (11)). Similarly to what discussed in [42], and consistently with the model in [43], we observe a power-law distribution in the exposure times (exponent $p < -2$), which emerges as a robust feature of random encounter dynamics. Additionally, we notice that the distance threshold plays a strong multiplicative effect and, possibly, it also weakly influences the exponent. It is worth remarking that our largest observation times are bound by the fact that we limit our graphs to time intervals of about 10-15 minutes around each train arrival. This reduces our resolution at large time scales and thus yields the exponential-like drop in the distribution tails.
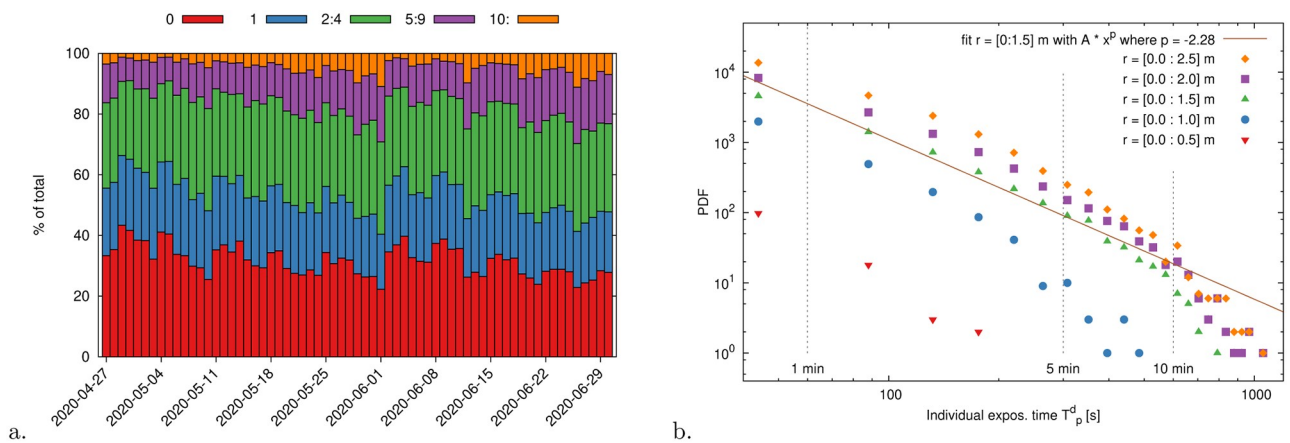
Fig 8. (a) Distribution of node-pedestrian degree per day as a percentage of the total number of passengers. The degree of a node counts the number of people encountered with a mutual distance smaller than 1.5m (hence, degree 0 means that a person did not have any Corona event). We observe that high-degree nodes, i.e. repeated distance offenders, increased steadily until the train schedule change (e.g. nodes with 10+ contacts grew from $\approx 1\%$ to $\approx 10\%$). The schedule change yielded a temporary drop in the offender percentage after which it started increasing again. This can be a sign of warning towards the relaxation in the compliance of physical distancing rules. (b) Probability density function of the individual exposure time discounted of families, $T_{p,f}^d$ considering different maximum distances (Eq (9)). Exposure times show a power-law behavior. The PDF depletion after $T = 5$minutes is most likely due to the time windowing that we operate around each train arrival (cf. Fig 4b). This yields a data cut-off for large times.
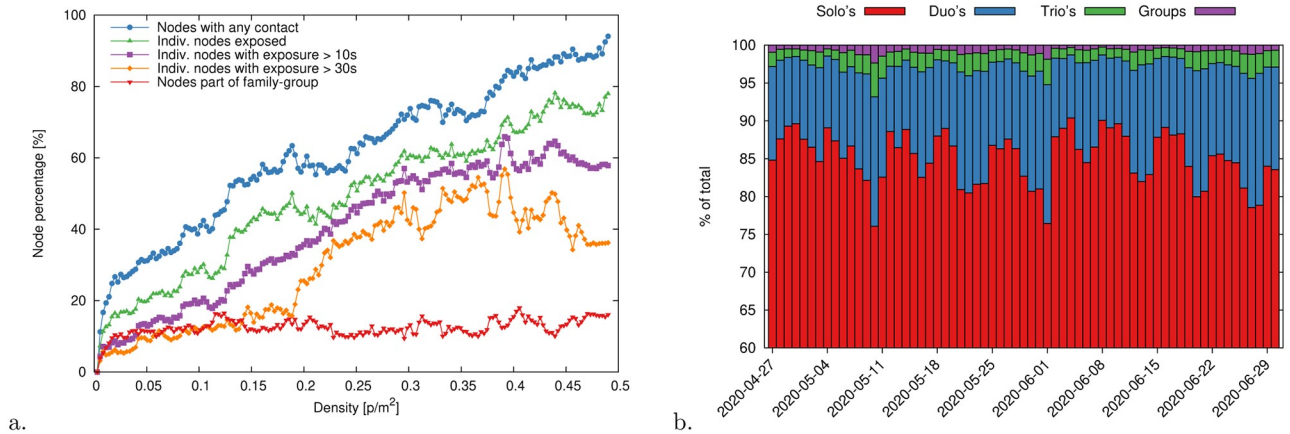
Fig 9. (a) Percentage of pedestrian nodes exposed to contacts as a function of the global density on the platform (density calculated as number of people in a frame divided by the total sensor area, 450m², discounted of the danger zone, 96m²). Exposed nodes that have at least one contact, of any duration, with another pedestrian (within or outside their family-group or not) are in blue. This percentage if further broken down into nodes part of a family group (red) and actual distance offenders (green). The purple and orange lines restrict, respectively, to nodes with a minimum exposure time of 10s and 30s. Linear fitting parameters are reported in Table 1. (b) Distribution of individuals and cliques day-by-day as a percentage of the total number of nodes. Between 80% and 85% of the nodes do not belong to cliques, i.e. they travel alone and their contacts are all distance infringements. Family-groups of two people cover about 12%–15% of the remaining nodes; family-groups of three and more provide a minimal ≈3% contribution.

https://doi.org/10.1371/journal.pone.0240963.g009

Crowd density: Incidence of family-groups and effective offenders

The passenger density on the platform has an influence on the offenses: the higher the density the easier it gets to violate physical distancing. In Fig 9a we report for a sample day (12th of June 2020), how the percentage of "family" nodes and offenders scales with the density within the global density interval [0, 0.5]ped/m². As in Eq (13), we also include minimum contact time thresholds, α, for tagging offenders. We observe that the percentage of nodes belonging to family-groups remains stationary (value ≈11%, a detailed breakdown of the clique size is in Fig 9b). Up to 80% of the pedestrians at the maximum density level committed offenses assuming no minimum time threshold ($\alpha = 0$). This percentage slightly diminishes to 60% restricting to a minimum contact time $\alpha = 5s$. Interestingly, the percentage shows a non-linear dependency on the density when $\alpha = 30s$. In particular, such percentage remains stationary (value ≈ 11%) until 0.2ped/m² and then suddenly increases. This can suggest an increase in difficulty in following distancing rules around this density level. We report the coefficients of the linear fitting of such data in Table 1.

Table 1. Linear fit parameters for the node classification (percentage data) in Fig 9a.

	$\delta \approx 0$	Slope
Ped.-nodes with any contact	28.5%	13.4%/δ
Ped.-nodes exposed	17.1%	12.9%/δ
Ped.-nodes with exposure $\alpha > 10s$	9.56%	12.4%/δ
Ped.-nodes part of a family-group	11.5%	0.48%/δ

As a measurement unit for the pedestrian density we employ *tenths of pedestrians per square meter*: $\delta = 10^{-1}$ped/m². Thus the fitting intercept is at $\delta = 0$ while the increments (slopes) are reported as percent variations per δ unit. As an example, pedestrians that do not belong to a family-group and are in contact with someone else grow about 13% when the density increases between 0.2 and 0.3ped/m² and similarly between 0.3 and 0.4 and so on. We do not report linear fitting parameters for the case $\alpha = 30s$ as the growth is non-linear.

https://doi.org/10.1371/journal.pone.0240963.t001

7 Discussion

We have presented an highly efficient and accurate approach to the problem of studying, in real-time, the distance-time encounter patterns in a crowd of individuals. Our approach allows us to identify social groups, such as families, by imposing thresholds on the distance-time contact patterns. In the context of the currently ongoing Covid-19 pandemic, we demonstrate this as an effective and promising tool to monitor, in a full privacy respectful way, the observation of physical distancing. The outcome of the analysis can provide early warnings in respect to an average relaxation towards the compliance of physical distance rules, can allow to identify spots where physical distancing most frequently is violated and it may, as well, allow to identify in real-time the presence of distance offenders. We observed, besides, a super-linear dependence between contact times and passenger number. This can be caused both by a reduction of attention towards social distancing rules but also to an intrinsic increase in difficulty in complying to regulations. The investigation of this aspect is left to future research.

The proposed algorithm is simple and can be easily implemented using existent graph code libraries. In our case, we could process a day of data in few minutes using the python NetworkX library [44]. Libraries sporting higher performance and/or scalability exist in case of even more demanding situations.

It is worth mentioning that the approach here proposed can be applied to any type of tracing trajectories and possibly to study the collective dynamics of large groups of active or passive particles making it a tool capable of going well beyond the application to crowd dynamics and physical distancing discussed here.

Author Contributions

Conceptualization: Federico Toschi, Alessandro Corbetta.

Data curation: Caspar A. S. Pouw, Alessandro Corbetta.

Funding acquisition: Frank van Schadewijk, Alessandro Corbetta.

Investigation: Caspar A. S. Pouw, Alessandro Corbetta.

Methodology: Alessandro Corbetta.

Project administration: Alessandro Corbetta.

Resources: Frank van Schadewijk, Alessandro Corbetta.

Software: Caspar A. S. Pouw, Alessandro Corbetta.

Supervision: Federico Toschi, Frank van Schadewijk, Alessandro Corbetta.

Validation: Alessandro Corbetta.

Visualization: Caspar A. S. Pouw.

Writing – original draft: Alessandro Corbetta.

Writing – review & editing: Federico Toschi, Alessandro Corbetta.

References

1. World Health Organization. COVID-19: physical distancing; 2020. Available from: https://www.who.int/westernpacific/emergencies/covid-19/information/physical-distancing.

2. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science. 2020; 368(6491):638. https://doi.org/10.1126/science.abb6105 PMID: 32234804

3. Rader B, Scarpino S, Nande A, Hill A, Reiner R, Pigott D, et al. Crowding and the epidemic intensity of COVID-19 transmission. medRxiv. 2020. https://doi.org/10.1101/2020.04.15.20064980

4. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet. 2020; 395(10228):931–934. https://doi.org/10.1016/S0140-6736(20)30567-5 PMID: 32164834

5. Chinazzi M, Davis J, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020; 368(6489):395–400. https://doi.org/10.1126/science.aba9757 PMID: 32144116

6. Sensormatic. Privacy-Preserving Contact Tracing; 2020. Available from: https://www.apple.com/covid19/contacttracing.

7. Cristiani E, Piccoli B, Tosin A. Multiscale Modeling of Pedestrian Dynamics. vol. 12 of Modeling, Simulation and Applications. Springer; 2014.

8. Helbing D. Traffic and related self-driven many-particle systems. Rev Mod Phys. 2001; 73(4):1067. https://doi.org/10.1103/RevModPhys.73.1067

9. Adrian J, Amos M, Baratchi M, Beermann M, Bode N, Boltes M, et al. A glossary for research on human crowd dynamics. Collective Dynamics. 2019; 4(A19):1–13. https://doi.org/10.17815/CD.2019.19

10. Moussaïd M, Perozo N, Garnier S, Helbing D, Theraulaz G. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. PloS one. 2010; 5(4):e10047. https://doi.org/10.1371/journal.pone.0010047 PMID: 20383280

11. Zanlungo F, Ikeda T, Kanda T. Potential for the dynamics of pedestrians in a socially interacting group. Phys Rev E. 2014; 89:012811. https://doi.org/10.1103/PhysRevE.89.012811

12. Gorrini A, Vizzari G, Bandini S. Age and group-driven pedestrian behaviour: from observations to simulations. Collective Dynamics. 2016; 1:1–16. https://doi.org/10.17815/CD.2016.3

13. Köster G, Seitz M, Treml F, Hartmann D, Klein W. On modelling the influence of group formations in a crowd. Contemporary Social Science. 2011; 6(3):397–414. https://doi.org/10.1080/21582041.2011.619867

14. Zanlungo F, Yücel Z, Kanda T. Social group behaviour of triads. Dependence on purpose and gender. Collective Dynamics. 2020; 5:118–125. https://doi.org/10.17815/CD.2020.41

15. Templeton A, Drury J, Philippides A. From mindless masses to small groups: conceptualizing collective behavior in crowd modeling. Review of General Psychology. 2015; 19(3):215–229. https://doi.org/10.1037/gpr0000032 PMID: 26388685

16. Brščić D, Zanlungo F, Kanda T. Density and Velocity Patterns during One Year of Pedestrian Tracking. Transportation Research Procedia. 2014; 2:77–86. https://doi.org/10.1016/j.trpro.2014.09.011

17. Corbetta A, Bruno L, Muntean A, Toschi F. High Statistics Measurements of Pedestrian Dynamics. Transportation Research Procedia. 2014; 2:96–104. https://doi.org/10.1016/j.trpro.2014.09.013

18. Corbetta A, Meeusen J, Lee C, Toschi F. Continuous measurements of real-life bidirectional pedestrian flows on a wide walkway. In: Pedestrian and Evacuation Dynamics 2016. University of Science and Technology of China press; 2016. p. 18–24.

19. Corbetta A, Meeusen JA, Lee C, Benzi R, Toschi F. Physics-based modeling and data representation of pairwise interactions among pedestrians. Phys Rev E. 2018; 98(6). https://doi.org/10.1103/PhysRevE.98.062310

20. Corbetta A, Lee C, Benzi R, Muntean A, Toschi F. Fluctuations around mean walking behaviours in diluted pedestrian flows. Phys Rev E. 2017; 95:032316. https://doi.org/10.1103/PhysRevE.95.032316 PMID: 28415258

21. Marchetti MC, Joanny JF, Ramaswamy S, Liverpool TB, Prost J, Rao M, et al. Hydrodynamics of soft active matter. Rev Mod Phys. 2013; 85:1143–1189. https://doi.org/10.1103/RevModPhys.85.1143

22. Brščić D, Kanda T, Ikeda T, Miyashita T. Person tracking in large public spaces using 3-D range sensors. IEEE Trans Human-Mach Syst. 2013; 43(6):522–534. https://doi.org/10.1109/THMS.2013.2283945

23. Seer S, Brändle N, Ratti C. Kinects and human kinetics: A new approach for studying pedestrian behavior. Transport Res C-Emer. 2014; 48:212–228. https://doi.org/10.1016/j.trc.2014.08.012

24. Kroneman W, Corbetta A, Toschi F. Accurate pedestrian localization in overhead depth images via Height-Augmented HOG. Collective Dynamics. 2020; 5:33–40. https://doi.org/10.17815/CD.2020.30

25. van den Heuvel J, Thurau J, Mendelin M, Schakenbos R, Ofwegen M, Hoogendoorn S. An Application of New Pedestrian Tracking Sensors for Evaluating Platform Safety Risks at Swiss and Dutch Train Stations. In: Hamdar S. (eds) Traffic and Granular Flow'17. TGF 2017. Springer; 2019. p. 277–286.

26. Giannotti F, Nanni M, Pinelli F, Pedreschi D. Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining; 2007. p. 330–339.

27. Zheng Y. Trajectory data mining: an overview. ACM Transactions on Intelligent Systems and Technology (TIST). 2015; 6(3):1–41. https://doi.org/10.1145/2743025

28. Zanlungo F, Brščić D, Kanda T. Pedestrian Group Behaviour Analysis under Different Density Conditions. Transportation Research Procedia. 2014; 2:149–158. https://doi.org/10.1016/j.trpro.2014.09.020

29. Benkert M, Gudmundsson J, Hübner F, Wolle T. Reporting flock patterns. Comput Geom. 2008; 41 (3):111–125. https://doi.org/10.1016/j.comgeo.2007.10.003

30. Vieira MR, Bakalov P, Tsotras VJ. On-line discovery of flock patterns in spatio-temporal data. In: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems; 2009. p. 286–295.

31. Kalnis P, Mamoulis N, Bakiras S. On discovering moving clusters in spatio-temporal data. In: International symposium on spatial and temporal databases. Springer; 2005. p. 364–381.

32. Bondy JA, Murty USR, et al. Graph theory with applications. vol. 290. Macmillan London; 1976.

33. Pitas I. Graph-based social media analysis. vol. 39. CRC Press; 2016.

34. Guo D, Liu S, Jin H. A graph-based approach to vehicle trajectory analysis. J Locat Based Serv. 2010; 4 (3-4):183–199. https://doi.org/10.1080/17489725.2010.537449

35. Silk MJ, Fisher DN. Understanding animal social structure: exponential random graph models in animal behaviour research. Anim Behav. 2017; 132:137–146. https://doi.org/10.1016/j.anbehav.2017.08.005

36. Ohtsuki H, Hauert C, Lieberman E, Nowak MA. A simple rule for the evolution of cooperation on graphs and social networks. Nature. 2006; 441(7092):502–505. https://doi.org/10.1038/nature04605 PMID: 16724065

37. Corbetta A, Lee C, Muntean A, Toschi F. Frame vs. Trajectory Analyses of Pedestrian Dynamics Asymmetries in a Staircase Landing. Collective Dynamics. 2017; 1(A10):A10. https://doi.org/10.17815/CD.2017.10

38. Hansen JP, McDonald IR. Theory of simple liquids. Elsevier; 1990.

39. Gallup AC, Hale JJ, Sumpter DJ, Garnier S, Kacelnik A, Krebs JR, et al. Visual attention and the acquisition of information in human crowds. Proc Natl Acad Sci USA. 2012; 109(19):7245–7250. https://doi.org/10.1073/pnas.1116141109 PMID: 22529369

40. Vallejo JC, Sanjuan MA, Sanjuán MA. Predictability of chaotic dynamics. Springer; 2017.

41. World Health Organization. Coronavirus disease (COVID-19) advice for the public; 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.

42. Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton JF, Vespignani A. Dynamics of person-to-person interactions from distributed RFID sensor networks. PloS one. 2010; 5(7):e11596. https://doi.org/10.1371/journal.pone.0011596 PMID: 20657651

43. Karagiannis T, Le Boudec JY, Vojnović M. Power law and exponential decay of intercontact times between mobile devices. IEEE T Mobile Comput. 2010; 9(10):1377–1390. https://doi.org/10.1109/TMC.2010.99

44. Aric A Hagberg DAS, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds); 2008. p. 11–15.