

# Mobility Timing for Agent Communities, a Cue for Advanced Connectionist Systems

Bruno Apolloni, Simone Bassis, Elena Pagani, Gian Paolo Rossi, and Lorenzo Valerio

**Abstract**—We introduce a *wait-and-chase* scheme that models the contact times between moving agents within a connectionist construct. The idea that elementary processors move within a network to get a proper position is borne out both by biological neurons in the brain morphogenesis and by agents within social networks. From the former, we take inspiration to devise a medium-term project for new artificial neural network training procedures where mobile neurons exchange data only when they are close to one another in a *proper* space (are in contact). From the latter, we accumulate mobility tracks experience. We focus on the preliminary step of characterizing the elapsed time between neuron contacts, which results from a spatial process fitting in the family of random processes with memory, where chasing neurons are stochastically driven by the goal of hitting target neurons. Thus, we add an unprecedented mobility model to the literature in the field, introducing a distribution law of the intercontact times that merges features of both negative exponential and Pareto distribution laws. We give a constructive description and implementation of our model, as well as a short analytical form whose parameters are suitably estimated in terms of confidence intervals from experimental data. Numerical experiments show the model and related inference tools to be sufficiently robust to cope with two main requisites for its exploitation in a neural network: the nonindependence of the observed intercontact times and the feasibility of the model inversion problem to infer suitable mobility parameters.

**Index Terms**—Algorithmic inference, brain morphogenesis, complex networks, mobile neurons, mobility models, Pareto distribution law, processes with memory, social networks.

## I. INTRODUCTION

A COUPLE of factors determining the success of complex biological neural networks, such as our brain, are represented by a suited mobility of neurons during the embrional stage along with a selective formation of synaptic connections out of growing axons [1]. While the second aspect has been variously considered in artificial neural networks, for instance in the ART algorithms [2] or simply through growing and pruning methods [3], a real mobility model of the single

neurons in the topological space where connections lie has not come into common use.

As for the biological side of this model, we may rely on convincing explanations of the mobility strategies leading to the definite brain morphology, like in the layered cortex [4], as well as on clinical evidence of diseases due to incorrect neuron migrations [5]. However, the neuron dynamics are related to chemical and electrochemical phenomena coupled with genetic primers [6] that prove difficult to handle quantitatively in a numerical simulation setup. Therefore, working jointly at different physical scales, we present in this paper a mobility paradigm that is both aimed at reproducing the main features of the epigenetic phenomena which characterize neuron mobility, and is checkable on a community of agents. As for the latter, we will refer to social networks as the most common implementation of the social computing paradigm [7], a set of persons who: 1) accomplish circumscribed tasks, but; 2) share a common interest; and 3) need a communication network to cultivate this interest [8].

Thus, we deal with three ensembles of computing agents: biological neural networks, at a micro scale, social networks at a macro one, and networks of artificial mobile neurons at an abstract level. The third represents the operational goal of this research, in the idea that mobility may constitute a further degree of freedom of the connectionist paradigm. This could for instance prove profitable to overcome the common drawbacks usually met in complex networks, such as the difficulty we have in training deep networks [9]. In this case, mobile neurons may find the proper reciprocal position by themselves, as do their biological counterparts in the various layers of the brain cortex. However, we focus on a very preliminary step of this medium-term goal. We simply elaborate on the neuron interaction timing and then compare our results with others present in the literature in analogous fields.

All agent ensembles share the connectionist paradigm according to which many elementary processors (agents in the role of, biological neurons, people, and artificial neurons) may develop a huge computational power, provided that the processors are highly interrelated and collectively driven toward a goal that may be transparent to the single processor [10]. The additional feature of neuron mobility intrinsically reflects epidemiological aspects of the inter-agent communication network. The leading idea is that, in an extended metric space taking into account both topological distances and functional distances ruled by the synaptic connections, only close neurons may efficiently interact with one another.

Manuscript received January 20, 2002; revised January 30, 2002; accepted September 8, 2011. Date of publication October 28, 2011; date of current version December 1, 2011.

B. Apolloni and S. Bassis are with the Department of Computer Science, University of Milan, Milan 20122, Italy (e-mail: apolloni@dsi.unimi.it; bassis@dsi.unimi.it).

E. Pagani and G. P. Rossi are with the Department of Information and Technology, University of Milan, Milan 20122, Italy (e-mail: pagani@dsi.unimi.it; rossi@dsi.unimi.it).

L. Valerio is with the Department of Mathematics, University of Milan, Milan 20122, Italy (e-mail: valerio@dsi.unimi.it).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNN.2011.2168536

Actually, the mobility mechanisms are definitely different in the three considered frameworks. However, we identify targeted trajectories out of variously expressed random walks as the common feature of the agent motion, which we reproduce with an abstract *wait-and-chase* scheme, that constitutes the core of our paper. It is a mobility model in the family of the stochastic point processes [11] which is unprecedented as to how it integrates Brownian motion [12], in the province of the waiting phase, with less symmetric paths denoting the chasing agent intentionality. Our focus is on the neuron dynamic linkage, which we characterize through the distribution law of the time elapsed between subsequent interactions. Then we devote a large part of our work to checking the model on real mobility trajectories of members of social networks. Both the mobility model and the operability of the emerging distribution law comprise the main results of this paper. As a further result, we show their relevance to the new envisaged connectionist paradigm in terms of a couple of statistical facilities complying with it. They are: 1) a set of outperforming inference tools allowing us to identify the distribution parameters from sequences of non independent interaction times (like in the artificial neuron mobility paths), and 2) a computationally affordable regression procedure as a solution to the problem of inverting the model, to infer the parameters of neuron dynamics which are synergistic with the training process of the neural network (we outline how in the conclusion section).

This paper is organized as follows. In Section II we introduce our wait-and-chase model. In Section III we discuss related research in the two reference frameworks of biological neural networks and social networks, and the new neural network concept as well. Then we move on to the experimental part. Namely, in Section IV we introduce the benchmarks we will use within the second framework. They are used to validate an inference procedure that we introduce in Section V to estimate the free parameters of the process. Then, in Section VI we regress these parameters on the mobility features of the model. In Section VII we compare these numerical performances with those of current models in the literature. Last, in the concluding section, we outline future work to train artificial networks of mobile neurons.

## II. MODEL TO MAINTAIN MEMORY IN A TIME PROCESS

Our focus is a (possibly) huge ensemble of particles moving in an Euclidean space of suitable dimension. A long-time favorite framework to study statistical properties of this ensemble is the random walk, where we consider the motion of each particle as a sequence of equal size steps taken in one of the directions allowed by the working space. At each step the direction is selected randomly, assigning equal probability each possible selection. This gives rise to Brownian motion trajectories [13]. Namely, thanks to the central limit theorem [14] after a large number of steps each trajectory coordinate in the Euclidean space follows a Gaussian distribution law that is centered on the starting point with a variance that linearly increases with the number of steps [see Fig. 1(a)]. The same theorem guarantees the reproducibility of these trajectories, i.e., the property that, by sequencing Brownian trajectories of

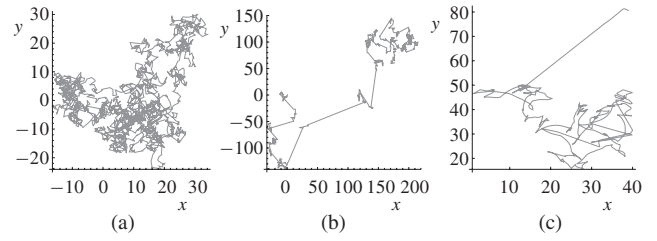


Fig. 1. Trajectories described by (a) Brownian motion, (b) Lévy flights, and (c) proposed mobility model.

any length and in any number, we obtain a new trajectory which is also described by a Gaussian distribution, yet with different parameters. This means that the motion emerging from an infinite sequencing of constant and randomly oriented steps is modeled as the infinite sum of independent Gaussian variables of 0 mean, constant variance on each direction, and zero covariance on each pair of orthogonal directions. For instance, these trajectories may describe the equilibrium motion of gas particles in a cylinder, or even the crowd at an expo, where the main trait is the full symmetry of their evolution.

On the contrary, a distinguishing feature of animate agents within a well structured ensemble - such as a social network - is the causality of their actions, i.e., their intentionality as the outcome of the agent memory. In turn, intentionality introduces local asymmetries in the motion which have been variously studied. A lead requirement is to maintain the mentioned reproducibility property. We get it by using Lévy flights [15] in place of random walks as elementary paths of the motion. A typical picture of a sequence of Lévy flights is shown in Fig. 1(b). It denotes paths which, besides even specific instances of human mobility [16], may describe the foraging patterns of animals such as albatrosses [17], namely, temporary stays in place (to eat and rest) plus sudden jumps here and there (to chase the food). We obtain a Lévy flight as a variant of the random walk for which an extended version of central limit theorem applies. This walk as well is defined as a sum of independent identically distributed random steps. The difference re Brownian motion is in the variance of the new steps which is normally not definite (broadly speaking, it is infinite) as a consequence of the above jumps. As a matter of fact, we may exactly define the trajectory coordinate distribution only in the Fourier transform space with uniquely asymptotical approximation in the original space [18].

By contrast, here we present a model having the same two-phase pattern as animal foraging, but with a greater prominence of local stays in place and therefore a less dispersed general behavior. This results in a bounded variance of the steps [see Fig. 1(c)]. The main benefit lies in a simple analytical form of the distribution law of relevant motion variables, at some expense of the reproducibility property. To introduce it, think of the dodgem cars at an amusement park.

### A. Very Abstract Framework and Its Model

Assume you are playing with dodgem cars. You drive around until, from time to time, you decide to bang into a

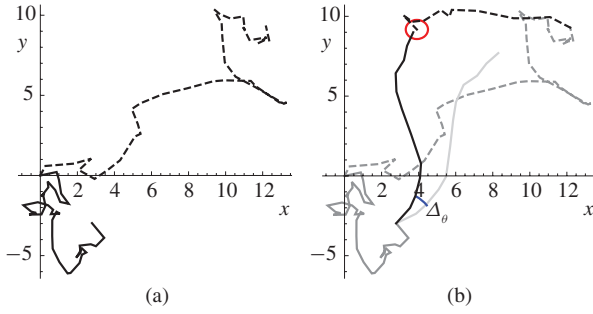


Fig. 2. Joint trajectories of two cars (plain and dashed curves, respectively) when (a) both move according to a Brownian motion and (b) former moves in a specific direction from a trigger time on, i.e., an oracle rotates this path by  $\Delta\theta$  toward the other car with some approximation (quantified by the ray of a proximity circle).

given car which is unaware of your intent. For the sake of simplicity, we may assume the trajectory of each car to be a plane Brownian motion before the chase triggering. Thus, with the reference frame in Fig. 2(a), indexing with  $i = 1, 2$  the cars whose stories we are following, we have

$$X_i(t) \sim \mathcal{N}_{0, \sqrt{t}}, \quad Y_i(t) \sim \mathcal{N}_{0, \sqrt{t}} \quad (1)$$

where  $\mathcal{N}_{\mu, \sigma}$  is a Gaussian variable of mean  $\mu$  and standard deviation  $\sigma$ . Then you, sitting in the first car, decide at time  $w$  to reach and crash into the second car. The questioned variable records the instant  $T > w$  when you succeed. In the case study, where cars are points in the plane, in order to identify this instant we must specify: 1) an operational definition of the cars' clash since the probability of exact matching is 0, and 2) the symmetry break introduced by the chase intention. The chase effectiveness depends on the ability to orient your motion in the direction of the target, which corresponds to converting a part of the motion along the cars' connecting line from symmetric to oriented moves. Mathematically, orientation corresponds to taking the chord of the cumulative trajectory in this direction [see Fig. 2(b)].

In order to overcome analytical complications and fulfill point 1) as well, we propose this simple scheme. As the difference between two Gaussian variables is a Gaussian variable too, we may use (1) also to describe the components of the vectorial distance  $\Delta$  between the two cars before  $w$ . We just need to multiply them by  $\sqrt{2}$  so as  $X_\Delta(t) \sim \mathcal{N}_{0, \sqrt{2t}}$  and similarly for  $Y_\Delta(t)$ . Moreover, if we move to polar coordinates  $(r, \theta)$  with  $x = r \cos \theta$  and  $y = r \sin \theta$ , the density function  $f_\Delta$  of  $\Delta$  becomes

$$f_\Delta(r, \theta) = \frac{1}{4\pi t} r e^{-\frac{r^2}{4t}} \quad (2)$$

which looks for the joint density function of  $(R, \Theta)$ , with  $R$  a Chi variable with 2 degrees of freedom scaled by a factor  $\sqrt{2t}$ , and  $\Theta$  a variable uniformly distributed in  $[0, 2\pi)$  independently of  $R$ . Our assumption about the pursuit is that, with reference to the distances  $D_1$  and  $D_2$  of the two cars from the position of the first one at time  $w$ , you are able to maneuver  $\Theta_1$  from  $w$  on, so that when  $D_1 = D_2$  also  $\Theta_1 = \Theta_2$  [see Fig. 2(c)].

Fulfilling point 2) with this assumption, we are interested in the time  $t$  when  $D_1 = D_2$ . Given the continuity of the latter we may measure only a probability density with  $t$ , which we may integrate in suitable intervals - corresponding to the proximity circle in Fig. 2 - to evaluate clashing probabilities. Since both  $D_1$  and  $D_2$  scale with the square root of time, expressing their dependence on the *trigger time*  $w$  and the *pursuit time*  $\tau$ , we have

$$D_1(\tau) = \sqrt{\tau} \chi_{21}, \quad D_2(\tau) = \sqrt{2w + \tau} \chi_{22} \quad (3)$$

where  $\chi_2$  denotes a Chi variable with 2 degrees of freedom whose density function is,  $f_{\chi_2}(z) = ze^{-(z^2/2)}$ . Thus, after equating  $D_1(\tau)$  with  $D_2(\tau)$  we obtain

$$1 = \frac{D_2(\tau)}{D_1(\tau)} = \frac{\chi_{22}}{\chi_{21}} \frac{\sqrt{2w + \tau}}{\sqrt{\tau}} \quad (4)$$

under the condition  $\chi_{21} \geq \chi_{22}$ . If we denote with  $\mathcal{T}$  the random variable with realizations  $\tau$  and  $W$  with realizations  $w$ , this equation finds a stochastic solution in the random variable

$$V = \frac{\mathcal{T}}{W} = 2 \left( \frac{\chi_{21}^2}{\chi_{22}^2} - 1 \right)^{-1}. \quad (5)$$

It follows the same distribution law of the ratio between two unconstrained Chi square variables, i.e., an  $F$  variable with parameters  $(2, 2)$  [19], whose cumulative distribution function (CDF) reads

$$F_V(v) = 1 - \frac{1}{1+v} I_{[0, \infty)}(v) \quad (6)$$

where  $I_{[a, b]}(x)$  is the indicator function of  $x$  w.r.t. the interval  $[a, b]$ , thus being 1 for  $a \leq x \leq b$ , 0, otherwise.

Since  $\tau = vw$ , to have the car pursuit time  $T$ , we need to have a convolution of the above distribution with the trigger time's. Let  $f_W$  be the probability density function of the latter, defined in a range  $(w_{\inf}, w_{\sup})$ . Since  $\tau + w = (v+1)w$ , we obtain  $F_T$ , with  $T = \mathcal{T} + W$ , by computing

$$F_T(t) = \int_{w_{\inf}}^{\min\{t, w_{\sup}\}} F_V\left(\frac{t}{w} - 1\right) f_W(w) dw. \quad (7)$$

The dependence of the convolution integral extreme on  $t$  induces a tangible dependence of the final distribution on the trigger time's. Nevertheless, keeping in mind some experimental results that we will discuss in the next sections, we look for a general shape of  $T$  distribution able to recover the mentioned dependences in a wide range of operational fields. To this end, we first generalize the form (6) into

$$F_V(v) = 1 - \frac{1}{1+v^{2\alpha}} I_{[0, \infty)}(v) g \quad (8)$$

obtained by changing  $\sqrt{t}$  into  $t^\alpha$  in (1). In this way we extend the scaling of the stochastic dynamics from the  $(1/2)$  power - used in the Brownian motion - to a generic power  $\alpha$  - in analogy to Lévy flights. Then, we approximate and further generalize this form through

$$F_V(v) = 1 - \frac{b+1}{b + \left(\frac{v}{c} + 1\right)^a} I_{[0, \infty)}(v) \quad (9)$$

with  $a, c > 0$  and  $b \geq 0$ , whose template shape is reported in Fig. 3 in terms of both  $F_T(t)$  in normal scale [see Fig. 3(a)],

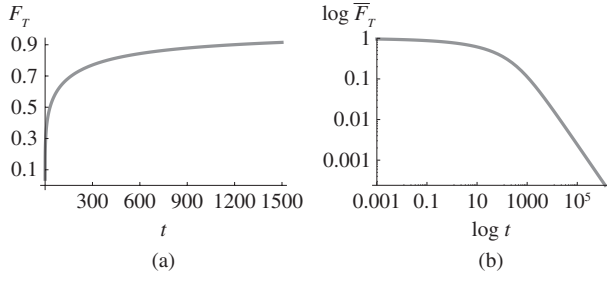


Fig. 3. (a) CDF plot of a shifted-Pareto distribution and (b) LogLogPlot representation of its complement.

and  $\bar{F}_T(t) = 1 - F_T(t)$ , i.e., the CCDF, in LogLog scale [thus representing both abscissas and ordinates in logarithmic scale, see Fig. 3(b)]. We call it a *shifted-Pareto* distribution since its typical elbow shape in the latter representation may be recovered from a basic Pareto distribution [20] just by shifting the time origin, i.e., through a law

$$\bar{F}_X(x) = \frac{b}{b + x^\alpha} I_{[0, \infty)}(x). \quad (10)$$

Though somewhat structurally different from (8) and (9) coincides exactly with (6) when  $a = c = 1$  and  $b = 0$ . Indeed, the main benefit we draw from (9) is the gain in model generality and flexibility, thanks to the three free parameters, we may get satisfactory approximations not only of (8) but also of (7) in a wide range of operational frameworks. Actually, by plugging (8) or (9) and the chosen trigger time distribution in (7), we obtain expressions whose analytical form is in general not easily computable. A better understanding of the interpolating role of (9) can be appreciated by referring to numerical integrations of (7), whose curves are reported in Fig. 4 for two typical instances. They refer to different trigger time distributions - emerging in slight changes in the length of the initial plateau, in the slope of the subsequent linear part, and also in the smoothness of their unions - which are well recovered by the parameters of (9).

Thus we will refer to the random variable  $T$  as a generic success time concerning the wait-and-chase process, expressed as a ratio  $V$  between chase  $\mathcal{T}$  and wait time  $W$ , or as pure chase time  $\mathcal{T}$ , otherwise as the sum  $\mathcal{T} + W$  of the two times. In any case, we expect this variable to be described by the CDF (9) with suitable parameters.

### B. Processes with Memory

The main trait of (9) is the power course of the CCDF with time - qualifying it as a heavy tailed distribution - in contrast to the exponential trend universally employed with temporal processes. We root it on the different dependence of the process on the memory backlog. In very essential terms, we speak of memory if we have a direction along which to order the events. Now, for any *ordered* variable  $T$ , such that events on their sorted values are of interest to us, the following *master equation* holds:

$$P(T > t | T > k) = P(T > q | T > k) P(T > t | T > q) \quad \forall k \leq q \leq t. \quad (11)$$

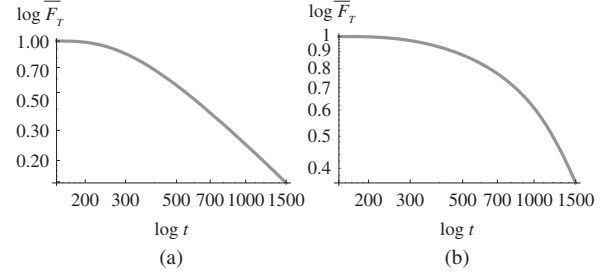


Fig. 4. CCDF LogLogPlot of contact times with a trigger time varying according to distribution law (a) Pareto and (b) uniform. Parameter  $\alpha = 0.9$  in (8) for all distributions, while parameters specific to the trigger distributions are set in order to have the same expected value.

It comes simply from the fact that in the expression of the conditional probability

$$P(T > t | T > k) = \frac{P(T > t)}{P(T > k)} = \frac{g(t)}{g(k)} \quad (12)$$

we may separate the conditioned variables from the conditioning ones. While (11) denotes the time splitting to be a general property of any sequence of data, (12) highlights that events  $(T > t)$  and  $(T > k)$  are by definition never independent. What is generally the target of the memory divide in random processes is the time  $t - k$  elapsing between two events. In this perspective, the template of the memoryless phenomena descriptor is the (homogeneous) Poisson process, whose essential property is  $P(T > t) = P(T > q)P(T > t - q)$ , if  $t > q$ . It says that if a random event (for instance a hard disk failure) did not occur before time  $q$  and you ask what will happen within time  $t$ , you must forget this former situation (it means that the disk did not become either more robust or weaker), since your true question concerns whether or not the event will occur at a time  $t - q$ . Hence your local variable is  $T - q$ , and the above property is satisfied by the (negative) exponential distribution law with

$$P(T > t) = \bar{F}_T(t) = e^{-\lambda t} \quad (13)$$

for constant  $\lambda > 0$ , since with this law (11) reads

$$e^{-\lambda(t-k)} = e^{-\lambda(q-k)} e^{-\lambda(t-q)} \quad (14)$$

and the property that  $g(t)/g(k)$  in (12) equals  $g(t-k)$  is owned only by the exponential function.

On the contrary, you introduce a memory of the past ( $q$ -long) if you cannot separate  $T - q$  from  $q$ . In this paper, we consider very simple cases where this occurs because the time dependence entails a local variable of the form  $(T/q)^\beta$ . The simplest solution of (11) is represented by

$$P(T > t | T > k) = \bar{F}_T(t) = \left(\frac{t}{k}\right)^{-\alpha} \quad (15)$$

so that the master equation reads

$$\left(\frac{t}{k}\right)^{-\alpha} = \left(\frac{t}{q}\right)^{-\alpha} \left(\frac{q}{k}\right)^{-\alpha}. \quad (16)$$

Equation (15) represents the basic form of the Pareto distribution [20]. It is defined only for  $t \geq k$ , with  $k > 0$  denoting the true time origin, and  $\alpha$  identifying the scale



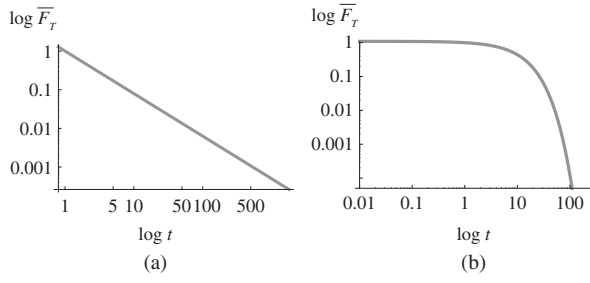


Fig. 5. CCDF LogLogPlot when  $T$  follows (a) Pareto law with  $\alpha = 1.1$  and  $k = 1$  and (b) exponential law with  $\lambda = 0.091$ . Parameters are chosen to have the same mean.

of its logarithm. The main difference w.r.t. the exponential distribution is highlighted by the LogLogPlots of  $\bar{F}_T$  in Fig. 5, a line segment with a Pareto curve [see Fig. 5(a)] in contrast to a more than linearly decreasing curve with the exponential distribution [Fig. 5(b)]. A first operational consequence is that, for a same mean value of the variable, we may expect its occurrence in a more delayed time if we maintain memory of it as a target to be achieved (getting a Pareto distribution), rather than relying on chance (getting an exponential distribution).

To link space and time aspects, we may consider Figs. 3(b) and 5(a) and (b) to be the time companions of Fig. 1(a)–(c) as representatives of a Brownian motion, Lévy flights, and our wait-and-chase process, respectively. In fact, it is well known that at sufficiently low densities the distribution of times between successive collisions of a molecule in a fluid is approximately exponential, while its trajectory follows a Brownian motion [21]. Analogously, experimental studies show a Pareto distribution reckoning the time intervals between changes in Lévy flight direction we see in nature. This occurs, for instance, with *Oxyrrhis marina* [22] or albatrosses [17] in search of food. In line with its space evolution, the trajectory timing of our process denotes an intermediate distribution between exponential and Pareto distribution. At a first glance we may identify in Fig. 3(b), an initial nonlinear plateau which we figure to have been drawn from the former distribution, followed by a linear part (henceforth denoted as the *tail*), referring to the Pareto distribution. This sharp characterization of the picture evinces the main features of our model. In Figs. 3(b) and 5(b), we enhance the contrast between the linear course of the tail of this distribution with respect to the more deeply decreasing course of the exponential distribution's by extending the ticks on the vertical axes toward values approaching much closer to 0. In Section VII we check the suitability of these features on experimental data, and contrast them with other models' considered in the literature.

### C. Completing the Mobility Model

Two additional points we must take into account to model the clash times are: 1) the interleaving of non intentional encounters with intentional ones planned by the chaser, and 2) the focus on the difference between subsequent clash times we are generally interested in as an instance of intercontact times in the mobility models. Non intentional encounters may be suitably framed into a Brownian process, so that both their

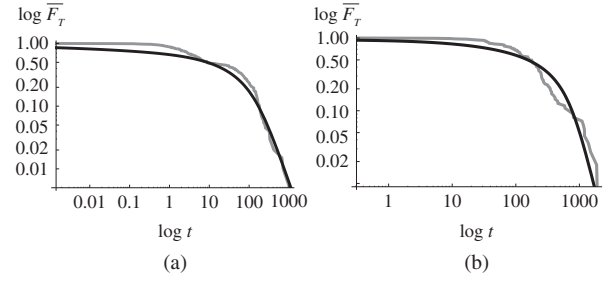


Fig. 6. Recovering the intercontact ECCDF shape with different mobility parameters through shifted-Pareto distributions with (a)  $a = 0.54, b = 11819, c = 1.27$ , and (b)  $a = 3.61, b = 7.55, c = 0.85$ . Gray curves: experimental distributions. Black curves: their interpolations.

absolute timing and the intercontact times (the local times of the process, see Section II-B) may be ruled by an exponential distribution law. By contrast, we have no strict theoretical results on the chaser intercontacts. However, relying on loose reproducibility properties deriving from both Brownian and Lévy processes, we use the same distribution law (9), with proper parameters, also for these differences.

We recover the two points numerically. For instance, denoting with  $\hat{\bar{F}}_T(t)$  the empirical complementary CDF (ECCDF) for a sample  $\mathbf{t} = \{t_1, \dots, t_m\}$  drawn from  $T$

$$\hat{\bar{F}}_T(t) = 1 - \frac{1}{m} \sum_{i=1}^m I_{(-\infty, t]}(t_i) \quad (17)$$

in Fig. 6 we see two ECCDFs LogLogPlot referring to a  $T$  sample obtained by drawing a trigger time uniform in  $[1, 150]$  and reckoning intervals between subsequent clash times whose ratio w.r.t. the former is simulated through (8). To these elements of the sample, we add further ones coming from an exponential distribution with a set parameter  $\lambda$ . In both figures we see a plateau analogous to those in Fig. 4, having the abscissa of its right end at around 100, followed by a linear slope that may be recovered through the shifted-Pareto distribution (9). The effect of the exponentially drawn times is the small hump on the sloping part that variously characterizes the experimental curves w.r.t. their interpolations through the mentioned distribution. Thus from this and similar graphs, we may recognize a general trait of the figures that we will use henceforth, where: 1) the length of the plateau plays the twofold role of order of magnitude of the mean trigger time, as for the constructive model, and of  $b^{(1/a)}$  in (9), as for the interpolating law, and 2) the slope of the tail is close to  $a$ . Continuing our broad characterization of the ECCDF shape to sharpen our intuition, we will refer to the first part of the curve as gathering the non intentional encounters almost exclusively, while to the second as describing almost entirely the intentional ones. On the contrary, the difference in the graphs in Fig. 6 is related both to the number  $m_e$  of exponentially drawn times w.r.t. the size of the sample connected to distribution (9), and to the values of parameters  $\lambda$  and  $a$ , respectively, in (8) and (13). These additional times have the dual effect of globally delaying the times within the plateau and of incrementing their number after its end. While in both cases  $\alpha = 0.9$ , the former has  $(m_e/m) = 0.5$  and

$\lambda = 0.1$ , the latter  $(m_e/m) = 0.4$  and  $\lambda = 0.0001$ . The corresponding interpolating parameters read as in the caption of Fig. 6.

### III. RELATED WORK

The model we introduced in the previous section has two main benefits: 1) it is a constructive model, so that we may discuss its main components from both physical and analytical perspectives in a clean albeit elementary framework, and 2) it is synthesized with a three-parameter formula, so that it allows for a relatively simple statistical identification. Clearly, it is a sharp schematization of a class of phenomena that usually occur with more complex features. In our paper we make the claim that this model is useful to synthesize mobility traits of the agents involved in the three - biological, social, and artificial - frameworks that we jointly consider in this paper.

As for the first, we distinguish two kinds of motion during brain morphogenesis, a physical one to migrate from the original neural cells production center (the neural plate) to their final position, and a virtual one where the sole axons and dendrites move to get efficient inter-neural connections. A typical template of the former is the aforementioned formation of the six layers of the brain cortex. The uniformity of the main parameters of the tissue, such as the number and kind of cells per volume unit, along the entire cortex may be referred to the genetic code of the single cells. However, the connection peculiarities of the various functional areas (such as vision, motion, etc.) in the cortex may be explained by epigenetic phenomena. A paradigmatic example of these phenomena is the growth cone dynamics through which axons of stably positioned neurons develop while seeking their synaptic target [23].

If we project them into a higher-dimensional topological space, both cell motion and axon growth are aimed at assessing mutual functional distances between neurons according to some efficiency criterion. *De facto*, we recognize the functionalities of the brain cortex's layered structure. Moreover, after a waiting phase where random walk features prevail, we physically observe the axon chase of the target cells, in spite of random obstacles that the axon cone must bypass along its trajectory. As for the specific rules of these dynamics, we know many motion mechanisms in terms for instance of chemotaxis [24], radial glia vehiculation [25], etc., [26]. However, their analytical synthesis in a unitary mathematical model - possibly comprehensive of the various side phenomena such as volume transmission, tunneling nano-tubes or receptor mosaics [27] - is still far from being realized. With our wait-and-chase model we operate the following trade-off. On the one hand, we expect to reproduce some traits of the two (random plus targeted) components of the neuron mobility in the high dimensional space. On the other hand, we abandon the idea of reproducing the specific mechanisms of the dynamics in favor of their statistical integration.

Who talk to us about the suitability of this modeling? As mentioned in the introduction, to gain a quantitative insight into it, we refer to some social network instances that are approachable in a metering space still preserving the main

features of the above biological framework. It is a common model transfer attitude in these fields, such as in the study of rat brain pharmacological reactions to infer analogous effects in people. In sum, we envision these instances as macro-scale instances of an improved connectionist paradigm deeply inspired by brain mechanisms.

While from a static perspective the main emphasis in social networks such as Facebook is placed on the relationship graph between the agents (the *role assignment* problem [28]), the time variability of these graphs is commonly dealt with in terms of dynamic network analysis through multiagent simulation techniques [29]. By contrast, we adopt a relatively recent approach to mobility which is connected to empowered telecommunication technologies, such as the mobile wireless *ad hoc* networks [30], and fully agrees with our connectionist framework. This approach to network evolution is rooted in the individual behavior of the single communicating members, having the network as a corollary rather than the reference gear of the dynamics. Epidemic processes [31] and opportunistic networks [32] are two instances of this focus which share with our biological/artificial neurons the following features: 1) the fact that the members of the network travel inside it; 2) they have a limited infection radius, so to get interactions one agent must come close to another; and 3) each member is primarily entitled to manage its mobility on its own. As for point a), we note that wired web communities [33] perform the virtual way of realizing mobility through connection link variations.

We find paradigms comparable to ours within a large variety of temporal dynamics having *complex networks* (in the recent acceptance [34]) as the common keyword. Thus wait-and-chase model represents an alternative approach with respect, for instance, to: 1) *rich gets richer* i.e., the preferential attachment [35] used both in economics [36] and in web consensus phenomena [33]; 2) transmission entropy minimization [37] for language analysis; 3) log returns in risk assessment for exceptional events in nature such as earthquakes [38]; and 4) analytically tractable mathematical models like the Markov chain of discrete sites for random waypoint processes [39] or geometric Brownian motion (GBM) [40] for income earning phenomena. We remark that, on the one hand, these approaches have theoretical roots which are totally different from ours. On the other hand, all the involved processes share the same intermediate position between random walk and Lévy flights that we highlighted in ours, so that the emerging distribution laws fall mostly within the generalized Pareto family [41]. In particular, with respect to the three-parameter formula proposed by Pareto himself over a century ago [42]

$$\bar{F}_T(t) = \left(1 + \frac{a(t-c)}{b}\right)^{-\frac{1}{a}} I_{[c,+\infty)}(t) \quad (18)$$

we introduce in (9), starting from (15), further scale parameters to render it adaptive to many experimental situations, but in a slight different way than in (18). This relatively eases the identification tasks required by our framework.

Wanting to formulate a connectionist model for the mobile agents in the two (biological and social) contexts we specifically consider, we may feed suitable dynamics to the neurons of an artificial neural network. This represents an additional

TABLE I  
DESCRIPTION OF REAL WORLD AND SYNTHETIC BENCHMARKS

Dataset		Location	Technology	# Agents	Mean Trace Length	# Processed Agents	Beaconing Time (sec.)
Artificial		Pentium IV Dual Core	—	39	275	39	1
PTR		Computer Science Dept. in Milan	PTR	39	1876	39	1.5
Crawdad	CH <sub>1</sub>	Intel Research in Cambridge	iMote	9	145	9	125
	CH <sub>2</sub>	Cambridge Univ. Computer Laboratory	iMote	12	342	12	125
	CH <sub>3</sub>	IEEE INFOCOM 2005 Conf. in Miami	iMote	41	510	41	125
HMTD	Orlando	Disney World	Garmin GPS 60CSx	37	68	18	30
	NCSU	Raleigh (North Carolina) Univ. Campus	Garmin GPS 60CSx	30	34	15	30
	NY	New York City	Garmin GPS 60CSx	35	25	18	30
	KAIST	Daejeon (Korea) Univ. Campus	Garmin GPS 60CSx	91	308	12	30
Nokia	Helsinki	Helsinki	GPS equipped mobile	522	212	50	1
	London	London	GPS equipped mobile	199	233	50	1

facility for exploring the parameter space, when we make the wandering of the neurons in the Euclidean space, where they physically lie, affect the search for the minima of the questioned cost function in the parameter space (in a way analogous to the simulated annealing *shaking*). A first consequence of this insertion is the relevance of the motion timing in the whole process. Though the training aspect is mostly rewarding for any theoretical or experimental effort, in this paper we opt to deal exactly with the timing of the network as a necessary preliminary step toward gaining the desired benefits. We are comforted in our effort by the fact that this is a descriptive task traversing many natural systems, from fluid dynamics [21] to micro-organism [22], animal [17], and human mobility [16]. This allows us to directly compare analytical forms and results.

We escape timing aspects connected with the neuron firing local mechanisms (see spiking neuron models [43]). Rather, the key assumption is that an exchange of information (say, forward propagation of the signal or backward propagation of the error in a layered structure) occurs only when the involved neurons are close enough to one another (i.e., are in contact). This entails two factors determining the connection strength between neurons: 1) the weight which accumulates corrections along the entire training history, and 2) a distance penalty factor which depends on the current position within the neuron trajectory. Also the dependence on the topological distance is not new *per se*. For instance, it is at the basis of the neighboring functions of the winner neuron in self-organizing maps [44]. However, here the penalty directly affects the information passing functionality of the neurons, which in turn determines the effective connection net. For the moment we do not hypothesize any Lagrangian of motion. Rather, we rely on the statistical rules described in Section II, which are triggered when an artificial neuron has something to exchange with others. Thus, in this case the *wait phase* of the atomic motion lasts until the neuron has something to transmit/receive and someone to/from which transmit/receive it, and the *chase phase* until it succeeds in transferring the message to the target companion. Plugging these motions into stationary dynamics, we re-obtain the distribution law (9) of the random variable reckoning the elapsed times between contacts stated within the moving neurons.

Though the interest in this distribution dates quite far back, our approach is innovative in that:

- 1) we explicitly deal with chaser intentionality in terms of a point process with memory, which lies in an intermediate position between random walks and Lévy flights;
- 2) we introduce a constructive model of a spatio-temporal process which: a) shares features with many phenomena in nature, and b) allows for bounded variance dynamics and clean analytical forms as well;
- 3) we add a new paradigm for stochastic processes which is quite different from those commonly supposed to root the above phenomena;
- 4) we add a new distribution to the family of generalized Pareto distribution laws which proves very suitable for the identification tasks required by our connectionist framework.

#### IV. REAL WORLD BENCHMARKS AND ARTIFICIAL DATASETS

The above discussion left us with a strong statement on a new family of connectionist paradigms, the efficacy of connection between artificial moving neurons depends at any given time on the topological distance between them, in addition to the connection weight. We may consider biological neurons to look for this efficiency during their reciprocal positioning and the growth of their neurites in the brain morphogenesis. Conversely, we might measure tracks of this mobility within the motion of agents in social networks which we assume to be looking for an analogous optimal positioning in order to exchange messages with desired interlocutors. We variously check these tracks for the suitability of our wait-and-chase model, using the distribution of intercontact times as our statistical touchstone for understanding the basics of the neuron mobility - the cue for future works. Since this task is not trivial, it will necessarily occupy most of the rest of this paper. However, it entails the assessment of some statistical tools of future use in our connectionist paradigm, as outlined in the final section.

We have gathered a huge amount of data on these intercontacts. Some comes from public or proprietary repositories

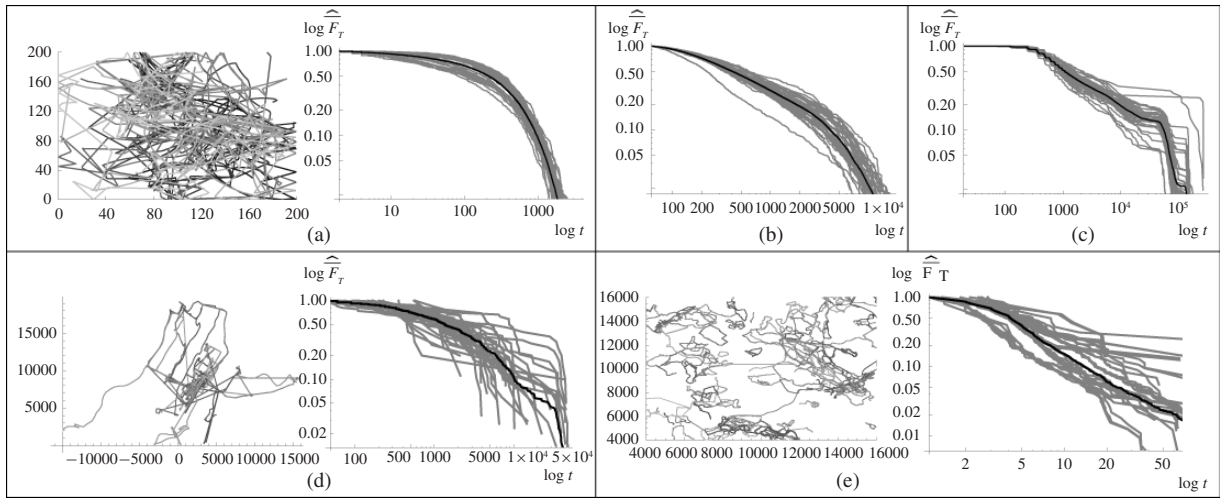


Fig. 7. Agent trajectories, wherever available, and tracks for some datasets, one per benchmark in Table I. Gray ECCDF curves: individual intercontact times, black ECCDF curves: merge of the former ones. (a) Artificial. (b) PTR. (c) Crawdad  $\mapsto$  CH<sub>3</sub>. (d) HMTD  $\mapsto$  New York. (e) Nokia  $\mapsto$  Helsinki.

available on the web. We have also produced our own experimental data to compare ours with the former's features, in addition we have simulated a huge dataset as an efficient *trait-d'union* between probability model and truth. The essentials of the benchmarks are reported in Table I, while a short description of them is reported here below.

#### A. Benchmarks from the Web

We check our methods on two categories of benchmarks: one expressly connected with opportunistic networks, the other to virtual communities which could implement analogous communication protocols thanks to contacts among members.

1) *Cambridge/Haggle (CH) Datasets*: They represent a collection of benchmarks available on the web, stored in the Crawdad database.<sup>1</sup> We focus on the three datasets (CH<sub>1</sub>, CH<sub>2</sub>, CH<sub>3</sub>) collected in the different research/academic sites listed in Table I. The data concern contact times between mobile agents endowed with an iMote platform [45] - an embedded bluetooth device equipped with a short-range antenna and limited battery (plus ARM processor and flash RAM). During the experiments, very short messages (*beacons*) are exchanged between the agents through the platform. Optimized algorithms deduce both contact - the period when two agents are in range of one another - and idle times - the period between the contact times, when data are not directly transferable between the two agents - through a beaconing strategy where a 5 seconds "enabled inquiry mode" alternates with a 120 seconds "sleep mode" (which accounts for a *beaconing time* of 125 seconds). We focus on pair-wise intercontact times, where the usual basic inspection tool is the ECCDF of the log of these times registered on a single individual getting in contact with any other participating in the experiment. We visualize the curves referring to the CH<sub>3</sub> dataset in LogLog scale in Fig. 7(c). Namely, the time logs constitute the *traces* of the single agents in the time domain as a companion of their *trajectories* in the space domain. The corresponding ECCDF

LogLog representations are the agent *tracks* from a statistical perspective. The trace *length* is measured in terms of number of reckoned intercontacts. In the following we will consider them both as a curve sheaf, one track for each agent (walker), and through single representatives. While when we refer to the interpolating parameters central values such as medians will be suitable, the overall shape of the tracks seems better preserved by the *merge* of the curves, i.e., the ECCDF of the merged traces.

2) *Global Positioning System Trajectories*: In recent years we have seen a wide proliferation of personal devices, mainly smartphones, endowed with GPS facilities. This has made plenty of mobility GPS trajectories available, as a result of current activities where agents may have different meeting modalities. To check extreme instances, we analyze two benchmarks where in the former - the human mobility trace dataset (HMTD) collected at the NC State University of Raleigh (NC)<sup>2</sup> encounters occur when people enter a relatively large interaction neighborhood re topological distances yet at the exact same time instant. We consider the complementary situation in the second benchmark, coming from Nokia's Sports Tracker project.<sup>3</sup> It concerns people who virtually interact with one another because they cross the same site although at (possibly) different times.

The trajectories in the HMTD dataset were collected worldwide on various cities/campuses through Garmin GPS devices. From the original signals, after a preprocessing phase (for instance isolating only the logs recorded within a radius of 10 km from the center of each site), and focusing on a single area where they were collected (e.g., New York City) we obtain the Cartesian coordinates of the walkers, as shown on the left in Fig. 7(d). Stating that a contact occurs when two people are fewer than 250 meters from each other, we obtain the companion intercontact times ECCDF as in Fig. 7(d) on the right. In Table I we distinguish between the

<sup>2</sup> Available at <http://research.csc.ncsu.edu/netsrv/?q=content/human-mobility-model-and-dtn-group>.

<sup>3</sup> Available at <http://sportstracker.nokia.com>.

<sup>1</sup> Available at <http://crawdad.cs.dartmouth.edu/cambridge/haggle>.



TABLE II

SYNOPSIS OF THE PARAMETERS FITTING THE BENCHMARK TRACKS. CELL VALUES: SINGLE TRACK COLUMN  $\mapsto$  MEDIAN AND MAD (IN BRACKETS) OF THE ESTIMATED PARAMETERS WITHIN THE DATASET, MERGE TRACK COLUMN  $\mapsto$  PARAMETERS FITTING THIS TRACK

		Single tracks				Merge track			
		$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$\tilde{d}$	$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$\tilde{d}$
Artificial		1.834 (0.215)	30584.6 (28461.6)	1.177 (0.046)	1. (0.)	2.97367	$4.46 \times 10^8$	1.02844	6
PTR		1.484 (0.115)	44920.6 (39575.6)	1.097 (0.042)	2.6 (0.6)	1.752	709781	0.959	4
HMTD	Orlando	1.018 (0.185)	31.881 (27.445)	1.261 (0.142)	0.2 (0.)	2.005	$2.08 \times 10^7$	0.959	8.
	NCSU	1.084 (0.262)	8336.36 (8334.31)	0.959 (0.388)	6. (2.)	1.329	151719.	1.531	6.6
	NewYork	1.099 (0.304)	10048.3 (9611.62)	1.133 (0.183)	2.6 (0.)	1.305	244650.	1.337	5.8
	KAIST	0.794 (0.17)	5.376 (4.204)	1.012 (0.21)	0.2 (0.)	1.819	$1.73 \times 10^7$	0.798	11.
Nokia	Helsinki <sub>dense</sub>	1.318 (0.324)	2.118 (1.517)	0.71 (0.067)	0.4 (0.2)	1.092	1.738	0.63	0.6
	Helsinki <sub>sparse</sub>	1.524 (0.334)	11.45 (10.881)	0.835 (0.094)	0.6 (0.4)	1.328	0.933	0.758	0.2
	London <sub>dense</sub>	1.883 (0.897)	32.886 (31.314)	0.863 (0.026)	1.2 (0.6)	1.646	1.985	0.829	0.2
	London <sub>sparse</sub>	2.922 (0.217)	715.551 (318.767)	0.869 (0.014)	1.7 (0.1)	2.991	1441.7	0.858	2.4
Crawdad	CH <sub>1</sub>	0.934 (0.166)	118.228 (98.693)	0.939 (0.128)	0.2 (0.)	0.879	106.163	0.926	0.3
	CH <sub>2</sub>	0.849 (0.122)	41.172 (26.771)	0.762 (0.111)	0.2 (0.)	0.977	530.91	0.724	0.8
	CH <sub>3</sub>	0.872 (0.082)	54.245 (32.943)	1.68 (0.24)	0.2 (0.)	0.813	31.488	1.725	0.2

number of available agents for each location and the number of them processed to get statistics. This denotes that some trajectories have been discarded in that tangibly anomalous (for instance, 2 or less intercontacts) w.r.t. the phenomenon we are considering.

The second dataset is the follow out of the Nokia Sports Tracker service (see Footnote 3) to which any person may apply by running a specific software on his own GPS-equipped smartphone. Now more than 125 000 trajectories are available, collected at a rate of one beacon per second from many regions around the globe. They drop to 9000 circa when we focus on people walking (neither running nor cycling or anything else) and to 522 and 199, respectively, in the cities of Helsinki and London, after bugged trajectories have been discarded. With reference to the pic on the left in Fig. 7(e), we isolated 236 trajectories spanning mainly at the bottom left corner. Then we variously sampled 50 tracks. Each one corresponds to a walker whom we jointly monitor with other (from 10 to 100) walkers whose trajectories we consider to reckon contacts. We did the same for the London tracks. In Table II we will distinguish between less crowded (sparse: fewer than 60 crossing walkers) and more crowded (dense) trajectories.

### B. Ground Truth

In order to get a clearer perspective of the phenomenon, we decided to essentially replicate the experiment with perfectly known environmental conditions. We achieved this both by developing portable radio devices (PTRs) and by deploying the test-bed.

We adopted the consolidated beaconing strategy in order to save batteries, with a beaconing time of 1.5 seconds. Whenever a beacon is received from a given encounter, the device creates a new entry in the local *contact-log*. The beacon contains the following items: 1) the local ID and ID of the encountered PTR; 2) the timestamp of the first contact; and 3) the timestamp of the contact closing. As for the latter, an entry in the contact-log is *closed* whenever no beacons

are received from the encountered device for more than 60 seconds. We collected data in two experimental campaigns performed between February and October 2008.<sup>4</sup> A set of 39 PTRs were distributed to students and administrative/teaching staff within the Computer Science Department of the University of Milano. At the conclusion of the campaign, their logs were remodulated so as to remove artifacts. In particular, we eliminated idle periods represented by the time intervals where people were expected to be away from the campus (between 7 PM and 8 AM. on workdays and all during the weekend). We also clamped to 0 the last 60 seconds of contacts artificially generated by the above connection closing rule. Tracks for all 39 agents are shown in Fig. 7(b).

### C. Artificial Dataset

We have numerically implemented the mobility model introduced in Section II on a simulated field. Namely, replacing the agents of the PTR campaign with dodgem cars, we consider a field of  $200 \times 200$  square meters with 39 individuals uniformly located inside it as our initial configuration. Each agent has two mobility modes, random waypoint [39] up to trigger time  $w$ , and the mentioned pursuit strategy after it. In the first mode, an agent randomly selects a direction that follows for a time length  $\theta$  uniformly drawn in  $[0, 2000]$  steps with a mean velocity of  $\bar{v} = 1.47$  meters per second (mean pedestrian velocity). This is simulated by tossing a positive random number less than or equal to 2000, as for  $\theta$ , a uniform value between 0 and  $2\pi$ , as for direction, and a random number drawn from a Chi distribution with 2 degrees of freedom scaled by  $1.17t$  (to maintain the mentioned mean velocity), to sample the distance  $D(t)$  covered by the agent at time  $t$ . At the completion of time  $\theta$ , it selects a new random direction and so on. When the trigger time  $w$  expires, it shifts to the second mode: the above Chi step is now coupled with a suitable angle rotation directing the agent toward the chosen target.

<sup>4</sup>Available at <http://nptlab.dico.unimi.it/index.php/cartoon.html>.

A match occurs when, for any reason, an agent gets closer than 10 meters to another one. We remark that we do not impose any constraint on the agent location (i.e., rebounds handling, etc.,) since the chase features automatically maintain an overall attractor within the original  $200 \times 200$  square. Fig. 7(a) reproduces a story of this model when the trigger time is drawn, for a mere choice of convenience, according to Pareto distribution (15) [see Fig. 4(a)]. In particular, dynamics parameters  $\alpha$  [the exponent in (8) modulating the agent mean speed versus time in this phase] and  $\nu$  [describing the rate of the trigger time distribution - as an instance of the exponent  $\alpha$  in (15)] were set to 0.9 and 1.5, respectively, and suitable values were chosen for ancillary parameters, such as  $k$  or the chased target distribution. The contact times are gathered along the entire experiment log corresponding to 18 days of simulated time.

## V. FITTING MOBILITY TRACES

The goal of the experimental part of our paper is to check the adequacy of our mobility model for interpreting mobility tracks observed in some agent communities. To this end, denoting with  $\mathbf{t} = \{t_1, \dots, t_m\}$  the sample of intercontact times of a given trace, we solve two orders of inference problems:

- 1) fitting of  $\mathbf{t}$  through form (9), by identifying *statistical parameters*  $a, b, c$  (*s-parameters*, henceforth) for each agent, with the aim of proving suitability of the proposed model to describe mobility;
- 2) regression of the *s-parameters* versus *mobility parameters* (for short *m-parameters*), which accounts for understanding the main traits of agent mobility. Here we focus on the parameters  $\alpha$  and  $\nu$  considered in Section IV-C, whose acquaintance allows us to better appreciate the departure from random walks, as we will show in the next section.

To be more convincing, we solve these problems in two steps, in terms of: 1) a reconstruction problem, by working with the artificial dataset introduced in Section IV-C, and 2) true inference problems over real world data described in Sections IV-A and B.

A first concern is to have an efficient automatic procedure to draw the interpolating curves. Actually, inferring a shifted-Pareto distribution is not a standard task *per se*. In addition, we must consider that, besides the hump discussed in Section II-C, empirical data are affected by many artifacts, linked for instance to seasonal phenomena such as user habits during a particular week and/or on a particular day of the week, special tasks shared exclusively by some pairs of users, etc., Thus, rather than expect a perfect fitting, we look for tight confidence regions where the experimental curves lie completely with a good probability. The identification of these regions is a favorite task of the Algorithmic inference paradigm [46] which we exploit here as follows.

### A. Statistical Bases

Looking at curves as in Fig. 7, we may consider our estimation problem in terms of drawing a regression curve through the set of pairs  $(t_i, \hat{F}_T(t_i))$ , coupling the observed intercontact

time with the ECCDF computed on it. According to our model, the regression curve depends on three s-parameters:  $a, b, c$ . In line with our stated goal, we look for a suitable region containing this curve, which we consider in principle as a realization of a random function. Thus, in analogy with the usual notion of confidence interval [19], we may define a *confidence region* as follows.

**Definition 1:** For sets  $\mathcal{X}, \mathcal{Y}$  and a random function  $C: \mathcal{X} \mapsto \mathcal{Y}$ , denote by abuse  $c \subseteq \mathcal{D}$  the inclusion of the set  $\{x, c(x); \forall x \in \mathcal{X}\}$  in  $\mathcal{D}$ . We define a confidence region at level  $\gamma$  to be a domain  $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$  such that

$$P(C \subseteq \mathcal{D}) = 1 - \gamma. \quad (19)$$

We infer this region via a bootstrap procedure in a slightly different version of the Efron paradigm [47]. The lead idea is that, starting from the observed data  $\{(t_i, \hat{F}_T(t_i))\}$ , we generate a huge set of curves that *could fit* them. They represent replicas of a random curve (i.e., a curve with random parameters) at the basis of these data, where the bootstrap generation method allows us to attribute a probability to each curve whose reckoning identifies the confidence region. With this perspective we devise a procedure running through the following steps.

1) *Sampling Mechanism:* It consists of a pair  $(\Psi, g_\theta)$ , where the *seed*  $\Psi$  is a random variable without unknown parameters, while the *explaining function*  $g_\theta$  is a function mapping from samples of  $\Psi$  to samples of the random variable  $X$  we are interested in [48]. Thanks to the probability integral transformation theorem [19] we have that, by using the uniform variable  $U$  in  $[0, 1]$  as a seed  $\Psi$ ,  $g_\theta$  coincides with the inverse of the  $X$  CDF  $F_X$ . In particular, the explaining function for  $T$  distributed according to (9) is

$$t = F_T^{-1}(u) = g_{a,b,c}(u) = c \left( \left( \frac{bu+1}{1-u} \right)^{\frac{1}{a}} - 1 \right). \quad (20)$$

2) *Master Equations:* The actual connection between the model and the observed data is exploited in terms of a set of relations between statistics on the data and unknown parameters that come as a corollary of the sampling mechanism. With these relations we may inspect the values of the parameters that could have generated a sample with the observed statistic from a particular setting of the seeds. Hence, if we draw seeds according to their known distribution - uniform in our case - we get a sample of compatible parameters in response. In order to ensure this sample clean properties, it is enough to involve sufficient statistics w.r.t. the parameters [19] in the master equations. Unluckily, because of the shift terms, the parameters are so embedded in the density function of  $T$  that we cannot identify such statistics for them. Rather we may rely on statistics that are *well behaving* (see [49] for their formal definition) with analogous benefits.

Namely, denoting by  $t_{(i)}$  the  $i$ -th element of the sorted intercontact times and by  $\overline{m}$  the quantity  $\lfloor (m+1)/2 \rfloor$ , we use the well behaving statistics

$$s_1 = t_{(\overline{m})}, \quad s_2 = \frac{1}{m} \sum_{i=1}^m t_i s_1, \quad s_3 = \sum_{i=\overline{m}}^m \log t_{(i)}. \quad (21)$$

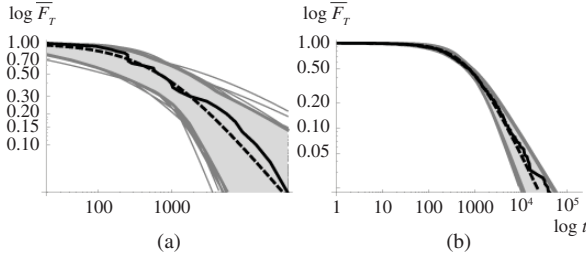


Fig. 8. Curves fitting with compatible parameters. (a) Sample size: 30. (b) Sample size: 500. Thick plain curves: sample ECCDF, gray curves: 200 population replicas, thick dashed curves: median of the replicas. Light gray region: 0.90 confidence region.

Thanks to the sampling mechanism (20) relating a realization of the uniform random variable  $U$  to a  $T$ 's, we obtain the master equations

$$s_1 = g_{a,b,c}(u(\overline{m})) \quad (22)$$

$$s_2 = \frac{1}{m} \sum_{i=1}^m g_{a,b,c}(u_i) - g_{a,b,c}(u(\overline{m})) \quad (23)$$

$$s_3 = \frac{\xi m}{2} \log c + \frac{1}{a} \sum_{i=\overline{m}}^m \log \left( \frac{bu_i + 1}{1 - u_i} \right). \quad (24)$$

As usual, we solve these master equations in the  $s$ -parameters in correspondence to a large set of randomly drawn seeds  $\{u_1, \dots, u_m\}$ . In this way, we obtain a sample of fitting curves, as in Fig. 8, which we statistically interpret to be *compatible* with the observed data. The two pictures differ only in the size of the sample generated through (20), sharing the same  $s$ -parameters  $a = 1.1$ ,  $b = 1000$ , and  $c = 1.2$ . The free parameter  $\xi$  is set to a value slightly greater than 1 in order to compensate the bias coming both from computing the last statistic only on a part of the observed sample, and, in the case of simulated/real tracks, from the truncation at the last intercontact, as a direct consequence of the finiteness of the campaign duration. In the figure, we also report the 0.90 confidence regions for these curves. We obtain these regions through a standard *peeling* method [50]. Namely, we circularly visited the upper and lower borders of the region made up of the envelope of the curves and erased the extremal ones, i.e., those that at least partially trespass the envelope of the remaining curves. We iterate the procedure until a fraction  $1 - \gamma$  of the original curves survives, where  $\gamma$  is the confidence level. Also, the thick line represents the median curve obtained with the same procedure, letting just one single curve survive (using obvious expedients when all the surviving curves, usually fewer than 5, lie on the border of the region). The pictures highlight the strong influence of the sample size on the width of the confidence region, which in any case contains the whole ECCDF uniformly over its support in both situations. We also note a sort of indeterminacy - to be read as non univocity - in the triples fitting the observed ECCDF. We may attribute this both to the high variance of the sample data and to the intricate interrelations among the trend of the curves and the  $s$ -parameters *per se*. Namely, though the large sample allows us to infer  $s$ -parameters closer to the

original ones ( $\check{a} = 1.14$ ,  $\check{b} = 1800$ , and  $\check{c} = 1.05$ , with  $\check{q}$  denoting the median estimate of parameter  $q$  using the above extreme peeling procedure), with the smaller sample we have acceptable interpolation as well, despite the great difference between the inferred parameters and their true values ( $\check{a} = 0.87$ ,  $\check{b} = 320$ , and  $\check{c} = 0.87$ ).

### B. Testing the Inference Algorithm

First of all we tested the procedure on a controlled environment represented by the artificial dataset introduced in Section IV-C. Fig. 9(a) completes Fig. 7(a) by adding the fitting curves obtained through our procedure. In greater detail, from the tracks of the single agents we get the confidence region at the top of the picture. Fitting intercontact times obtained by merging individual trajectories, we get the dashed curve at the bottom of the picture which proves very close to the merge ECCDF curve.

We did not have the same success with real tracks. This motivated us to do a deeper analysis of the data described in Sections IV-A and IV-B, suggesting that the hypothesis of their independence should be removed. While with intercontact times derived from simulation this hypothesis is true (within the limit of our pseudo-random number generation), with real people we may expect non independent performances. So, we may expect that with busy, sociable and/or even anxious people, if a task required a quick contact with colleagues, a subsequent one will do the same with a high probability, idem for encounters that do not last very long. On the contrary, a different attitude may induce a greater duration after a rushed task and *vice versa*. In our context, this essentially calls for, respectively, positive and negative autocorrelation among intercontact times.

Inference methods from correlated data represent a big issue in statistics, for which a vast variety of solutions has been offered, with most depending on the features of the random process that generates the data. Time series analysis in econometrics processes is a well-assessed vein in this field, having in [51] a groundbreaking text. Far from enunciating a further general theory on non-independent samples, here we simply consider an expedient to extend our procedure to this kind of data. Namely, we reverberate the dependence among times directly on the observation seeds  $u_i$ s entering the sample mechanism (20). We may look directly at a Markov process on the seeds, so that their conditional distribution depends on certain batches of previous seeds. If we are so lucky to have some standard conditions satisfied [52], we may rely on an equilibrium distribution from which to pick more suitable seeds of the sampling mechanism of the observed non independent intercontact times. Using the special typed symbol  $U$  (resp.  $u$ ) to distinguish the new seed from the uniform variable  $U$  (or its realization  $u$ ), we have a very preliminary hypothesis on its CDF as follows:

$$F_U(u) = u^{\frac{1}{d}} \quad (25)$$

with  $d > 0$ . It is definitely a gross hypothesis, relying simply on some similitude between the ECDF of samples generated by the mechanism  $u_i \equiv u_i^d$  [hence from the random variable

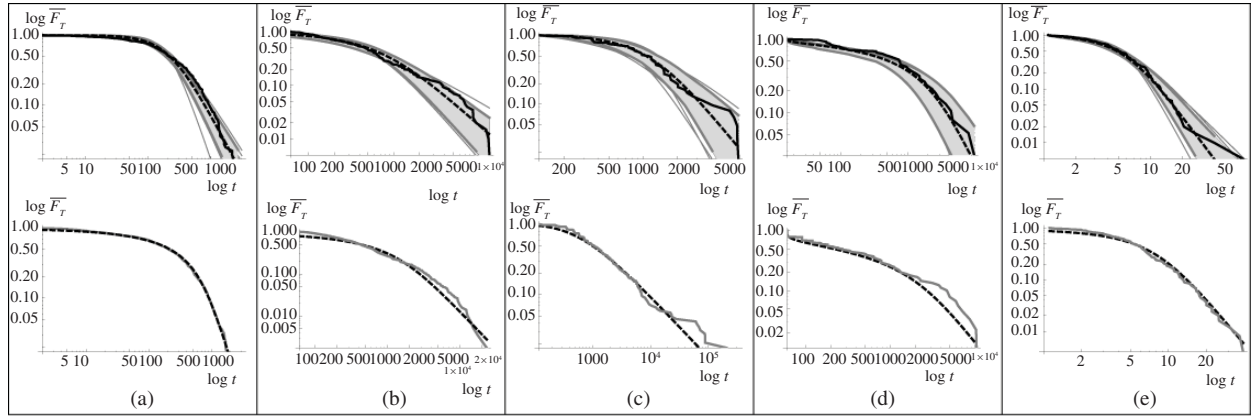


Fig. 9. Fitting agent tracks drawn from the dataset in Table I through our shifted-Pareto distribution. First row: 0.90 confidence region and median curve for single agents, same notation as in Fig. 8. Second row: merge track of all agents from the same dataset (gray curve) and its fit (black dashed curve). (a) Artificial. (b) PTR. (c) Crawdad  $\mapsto$  CH<sub>1</sub>. (d) HMTD  $\mapsto$  Orlando. (e) Nokia  $\mapsto$  London.

$U^d$  having exactly the CDF (25)] and the sampling mechanism  $u_i \equiv u_i^{(u_{i-1}/h)^r}$  (reproducing a Markovian dependence of the current  $u_i$  from the previous one  $u_{i-1}$ ), for suitable  $d$  as a function of  $r$  and proper tuning parameter  $h$  (see Fig. 10). As for the autocorrelation  $\rho_{U_i, U_{i+1}}$ , the value  $d = 1$  denotes independence between sample items, whereas  $d < 1$  corresponds to  $r > 0$  and  $\rho_{U_i, U_{i+1}} > 0$ , and  $d > 1$  to  $r < 0$  and  $\rho_{U_i, U_{i+1}} < 0$ .

On the one hand, with known  $d$  nothing changes on the above statistical procedures, apart from the new seed generation, since the sampling mechanism now reads

$$t = c \left( \left( \frac{bu^d + 1}{1 - u^d} \right)^{\frac{1}{a}} - 1 \right) \quad (26)$$

leading to the CDF

$$F_T(t) = \left( 1 - \frac{(b+1)}{b + \left( \frac{t}{c} + 1 \right)^a} \right)^{\frac{1}{d}}. \quad (27)$$

On the other hand, adding  $d$  within the set of  $s$ -parameters would unbearably complicate the procedure. Thus we decide to use  $d$  as an external parameter that we infer by inspection by maximizing the correlation  $\rho_{T, \tilde{T}}$  between actually observed times  $t_i$ s and reconstructed times  $\tilde{t}_i$ s. The latter are obtained by inverting the inferred CDF  $F_{\tilde{T}}$  on the ECDF  $\hat{F}_T$  specifications computed on  $t_i$ s. Namely

$$\tilde{t}_i = (t : F_{\tilde{T}}(t) = \hat{F}_T(t_i)). \quad (28)$$

Hence we relate  $t_i$  to the  $(i/m)$ -th quantile  $\tilde{t}_i$  of  $F_{\tilde{T}}$ , with  $i = 1, \dots, m$ .

We thus obtain the other pictures in Fig. 9, which complete the analogous ones in Fig. 7, referring to different datasets within the same benchmarks, so as to broaden our inspection of the data. We note that the confidence regions satisfactorily include the empirical curves, whereas the median of the compatible curves absorbs the hump commonly occurring over the tails. As discussed earlier, we may attribute this to a superposition of casual encounters, which parallel the intentional ones. While the bending of the ECCDF around the

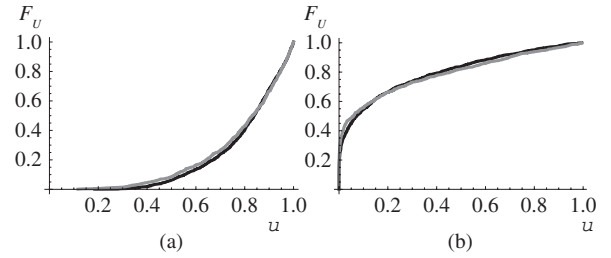


Fig. 10. ECDF of samples drawn according to the sampling mechanism  $u_i \equiv u_i^{(u_{i-1}/h)^r}$  (gray curves) and  $u_i \equiv u_i^d$  (black curves) when (a)  $r = -1$ ,  $h = 1$ ,  $d = 4$ ,  $\rho_{U_i, U_{i+1}} = -0.24$  and (b)  $r = 1$ ,  $h = 3$ ,  $d = 0.25$ ,  $\rho_{U_i, U_{i+1}} = 0.37$ .

estimated median curve when using the Crawdad and HMTD benchmarks hints at the presence of further local phenomena that generate these balanced shifts, the more regular course of the Nokia tracks may depend on its different method of gathering intercontact data. With the former benchmarks we reckon the time difference between one contact and the next for one agent versus another specific one for each paired peers, and subsequently gather the differences referring to the same agent. With the latter benchmark, we reckon the time difference between one given agent's contacts with any other agent. Regarding the merge curves, again we observe a good fitting of the inferred median parameters.

### C. Overall Evaluation

In Table II, we sum up the above inference on the benchmarks listed in Table I. Our main statistic is the median of the parameters of the compatible curves computed for each agent. In turn, of these values we report in the first column the median and the deviation around it in terms of median absolute deviation (MAD) [53], to capture respectively the central trend and the dispersion over the agents. In the second column, we refer directly to the merge traces of the various benchmarks for which we analogously report the compatible curve median parameters. A first insight arising from an overall outlook is that with our model we cover a vast variety of targeted short-range walking situations, w.r.t. walking mode (relaxed,

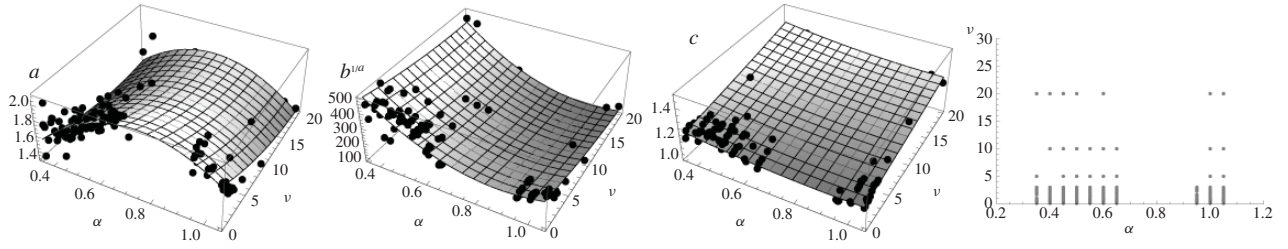


Fig. 11. Relation between s-parameters and m-parameters in the artificial dataset. Surfaces: best fitting curves, points: s- and m-parameters.

jogging), common goals (work, entertainment, sport, shopping, etc.), geographic location (from Europe to U.S.A. or Asia), and recording method (merge of specific pair contacts, one trajectory crossing a variable number of other trajectories, merge tracks). This reverberates in a parameter variety, albeit with some structural commonalities.

From a modeling perspective, we note that the weak reproducibility property of our shifted-Pareto distribution hypothesized in Section II-C emerges in the parameters as well (with some exceptions). Indeed, taking note that the plateau parameter to be compared is  $b^{1/a}$ , we find the parameters of the merge track close enough to the medians of the single agent parameters. The main exceptions occur for Orlando and KAIST), while minor discrepancies arise on a restricted number of parameters in three other sites. There is no clear relationship between these discrepancies and the dispersion of the parameters drawn from the single agents. Rather, we note that they never occur alone, but on at least a couple of parameters per benchmark. Re the variability of the solutions of the inversion problem (22)–(24) we mentioned in Section V-A and with the chimeric effects due to track mixing as well, the above fact might suggest attributing discrepancies to numerical artifacts rather than to statistical reasons. In this sense, paradigmatic is the correlation described by parameter  $d$ . This parameter takes values much greater than 1 in the merge curve of the artificial benchmark, despite the true intercontact independence (by construction) in the single agent traces.

As for the parameter dispersion, which we appreciate through MAD, we actually do not expect moderate values as an indicator of good estimates. Rather, on the one hand, their high values are again a consequence of the different experimental conditions each track refers to, where the near coincidence with median for  $b$  denotes an exponential distribution law of this parameter. On the other hand, we cumulated a set of approximately 300 traces for a total of 140 000 circa observed intercontacts that we will exploit as a whole in Section VII to have very meaningful fitness tests. For the moment, we get rid of the experiment variations through the structural coherence between data and use the different values as a cue to characterize the experiments.

## VI. UNDERSTANDING THE MOBILITY MODEL

In the previous section, we were left with a satisfactory match between experimental data and their modelization. Now we want to exploit this favorable condition to invert the model, i.e., to deduce the m-parameters from the s-parameters. It accounts for a regression problem that we solve using artificial

datasets like in Section IV-C as training set. Namely, we kept  $\alpha$  and  $\nu$  exponents as free parameters in a range compliant with the PTR setup. In detail, we let  $\alpha$  vary in the range  $[0.35, 1.05]$ , so as to appreciate sensible deviations from the Brownian motion setting (i.e.,  $\alpha = 0.5$ ), and  $\nu$  in  $[0.1, 20]$ , spanning a wide range of Pareto rates to cover both finite and non finite moments of this distribution (see rightmost picture in Fig. 11). Note that the  $d$  parameter is out of the question, since it is constantly equal to 1 in the artificial setup. To learn the regression function of the remaining s-parameters  $a, b, c$  versus m-parameters  $\alpha$  and  $\nu$ , first we identify the median as a template of the CCDF curves, then we regress its parameters through a polynomial in the m-parameters.

In Fig. 11 we see the best fitting that we obtain separately on  $a$ ,  $b^{1/a}$ , and  $c$ . The interpretation of these curves is far from simple. Here we just venture some guesses, declaring in advance that they are partial and need serious validation. We note that the first graph shows a complex trend of  $a$  with  $\alpha$  that we interpret as follows. On the one hand, the analysis of the course of  $\log \bar{F}_T$  with  $t$  shows that an  $a$  increase ( $a \uparrow$ ) in (9) has the effect of shifting the elbow between the non-Pareto and Pareto parts back (as for turning time) and down (as for the corresponding probability). This produces the twofold effect of reducing both the distribution time scale ( $t \downarrow$ ) and the rate of contact times ( $r \downarrow$ ) falling in the second part (call them the intentional times according to our broad curve interpretation in Section II-C). On the other hand, we see that  $a$  has a parabolic trend with  $\alpha$  having the top in the surrounding of  $\alpha \approx 0.5$ , a value that calls for the Brownian motion as the basic component of the model. Moving far from this value, we see a decreasing of  $a$  that we alternatively relate to the two effects  $t \downarrow$  and  $r \downarrow$ . Namely, since  $\alpha$  is a mobility speed-up factor, on the left-hand side of the trend we relate the increase of  $a$  with  $\alpha$  to a decrease in the time scale ( $t \downarrow$ ). This effect is contrasted by the rarefaction of the random encounters when  $\alpha$  becomes still higher, since the probability of crossing a same 10 meter raw transmitting coverage diminishes with the velocity due to the low agent density. Under these conditions, we have an overwhelming amount of intentional contacts (belonging to the Pareto part) ( $r \uparrow$ ).

We may similarly explain the second graph, where we broadly relate the  $b^{1/a}$  parameter to the scale of the non intentional encounter times. In principle, this scale decreases with  $\nu$  - since the average of the related Pareto does so - and increases with  $\alpha$  - because of the aforementioned spreading effects of this parameter. However, in this case too we have a saturation effect, so that very small  $\nu$ s equalize the trigger



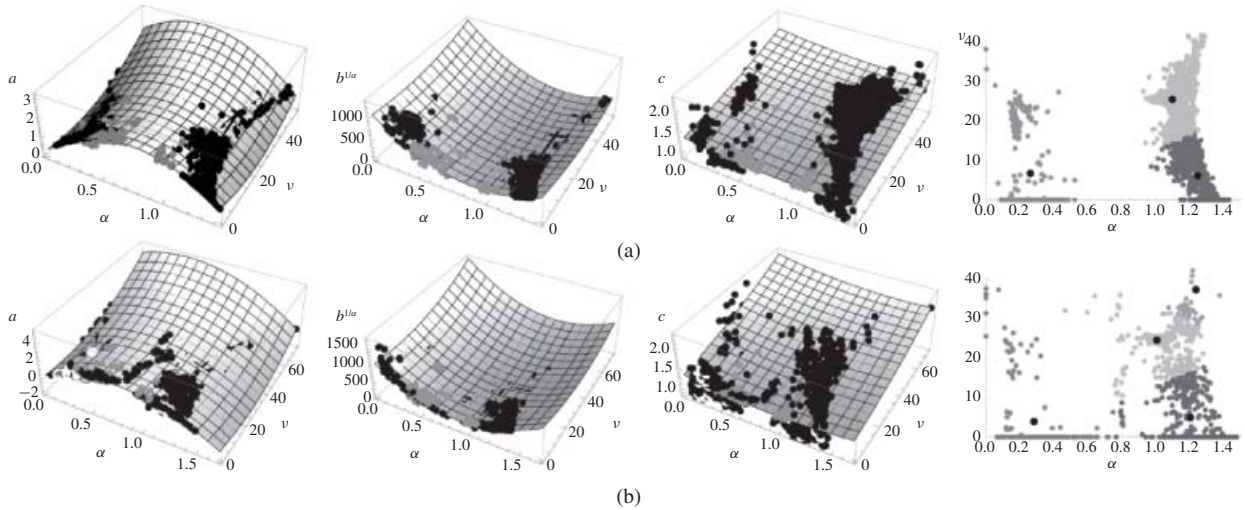


Fig. 12. Mobility parameters emerging from smearing the s-parameters of our experimental benchmark tracks on the surfaces in Fig. 11. First three columns  $\rightarrow$  gray points: same simulated parameters as in Fig. 11, black points: replicas compatible with the processed dataset, white points: median parameters among the replicas (graphically hidden in general by the former ones). Last column  $\rightarrow$  gray points: mobility parameters colored according to the cluster they belong to, bullets: cluster centroids. (a) Crawdad. (b) HMTD.

times. As a consequence, the number of (now almost purely intentional) contacts follows a Poisson distribution that is analogous to the one of the almost purely non intentional encounters reckoned in the opposite corner. Likewise, we have seen that short  $a$ s in correspondence to short  $a$ s may reduce the number of non intentional encounters (since  $r \uparrow$ ) contributing to the definition of the scale of the non Pareto part.

The third parameter,  $c$ , looks like a fine tuning factor indirectly affected by the m-parameters.

Moving to the experimental datasets, we want to discover through the above regression curves both the mean velocity and the mean waiting time of the people who wear the beaconing device. Namely, having computed s-parameter replicas compatible with the collected experimental datasets through master equations (22)–(24), as in Section V, we look for the corresponding m-parameters  $a$  and  $\nu$  that minimize the relative error between the computed  $a, b, c$  and the triple obtained through the regression curves. We processed all benchmarks in this way. However, in Fig. 12 we replicate Fig. 11 only on two of them (merged re benchmark subsets) for space reasons, claiming similar features for the unrepresented pictures.

We see that the first two columns denote a notable generalization of the regression curves. The clouds of points refer to the union of around 30 000 curves (hence the triplets of parameters specifying them) that are compatible with the single agent tracks. For all these curves, on the one hand we obtain values in line with the overall trend of both  $a$  and  $b$  with  $\alpha$  and  $\nu$ , as modeled in the previous sections. On the other hand, these values are compatible with the velocity and idle-time features of the people running in the various experiments and reflect the two polarizations of the dynamics (before and after  $\alpha = 0.5$ ) discussed in the previous section.

With  $c$  things go worse. But this is to be expected given the tuning role of this parameter. We note, however, that taking into account its value in back-regressing  $\alpha$  and  $\nu$  (through the minimizing procedure) diminishes the spread of these parameters.

The fourth column in Fig. 12 highlights the great similarity between the mobility parameters underlying the different benchmarks. This comes through in the shape of the clouds gathering them and even in the location of the centroids of the clusters emerging from an elementary k-means algorithm computed on the whitened data to take into account the different dispersion of the mobility features [54].

## VII. COMPARISONS WITH THE LITERATURE

The puzzling point of the analysis of the above data is that a certain number of distribution laws have been considered in the literature to this aim, all of which have found a tangible number of promoters. Some distributions emerge as a corollary of different theoretical approaches, like those mentioned in Section III, others come simply from a clever assembling of distribution features that fit the observed data. Coming to our benchmarks, we realize that the common features emerging from the various tracks include: a prominent elbow, which separates the plateau from the slope, and a linearity of the latter (the tail). Actually, the elbow is an artifact of the CCDF LogLog representation of a vast variety of distribution laws, from uniform distribution to Gaussian, and even to the exponential one. Among them, the following distributions are adopted in the literature as candidates for coping with a prominence of this feature: exponential, lognormal, tapered Pareto, and truncated Pareto. Thus, in this section we will employ these as competitors of our shifted-Pareto distribution. An exception w.r.t. this feature is precisely represented by the Pareto distribution - in essence a power law - whose graph, on the contrary, fully meets the second feature of our tracks (Figs. 7 and 9).

In linear scales, we analytically describe the plateau with a nearly exponential course and the tail with a definite power law. Wanting to explicitly maintain the dichotomy, other authors have given different readings of these traits. We might distinguish two gross families of such distributions: 1) a *double Pareto* curve (a lower power curve followed by a greater

TABLE III  
STATISTICAL COMPARISON BETWEEN COMPETITOR MODELS. ROWS: BENCHMARKS, COLUMN: MODELS, CELLS: CvM TEST  
ACCEPTANCE RATE (UPPER LINE) AND AKAIKE CRITERION WINNING RATE (LOWER LINE)

	Shifted-Pareto Best	Shifted-Pareto Median	Exponential	Pareto	LogNormal	Tapered Pareto	Truncated Pareto
PTR	0.154	0.077	0.821	0.	0.026	0.	0.077
	0.949	0.923	0.	0.	0.026	0.	0.
Crawdad	0.508	0.169	0.067	0.051	0.034	0.017	0.068
	0.441	0.101	0.	0.	0.288	0.	0.271
HMTD	0.593	0.468	0.	0.25	0.468	0.562	0.781
	0.656	0.469	0.	0.	0.	0.062	0.281
Nokia	0.937	0.796	0.	0.312	0.641	0.437	0.328
	0.531	0.281	0.	0.218	0.141	0.	0.109

power one) [40], [55], and 2) a temporal sequencing of a Pareto course proceeding with an exponential distribution that quickly nears 0 [56].

The former is analytically derived by the GBM model [40]. However, though extensively studied in growth phenomena [36], [57], it has been adapted to mobility studies with feeble results. In particular, it misses the real course of the first part of the experimental curves. *Vice versa*, the tapered Pareto distribution [58]

$$\bar{F}_T(t) = 1 - \left(\frac{k}{t}\right)^\alpha e^{\frac{k-t}{b}} I_{[k,+\infty)}(t) \quad (29)$$

and similar ones (namely lognormal, truncated Pareto) [59], [60] directly structured to fit in the second alternative, have the exponential decrease of the function in the rightmost part of the curve as their main drawback. This is somehow explained in terms of nomadic motion [61], and finds a good fitting only with specially devised traces [62]. As a matter of fact, researchers working on Nokia datasets also [63] lean toward an analytical description of these data through a tapered Pareto distribution, though admit that other types of mobility patterns, and consequent distributions, could equally serve their purpose.

Other authors prefer concentrating their analysis on the most populated part of the traces to gain simplicity. Thus they bind the analysis near the plateau, lowering the relevance of the remaining times with the twofold effect of shading times that are exceedingly long and extremely costly to process re parameter estimation [56], [64]. Then, they analyze the surviving data according to the other candidate distributions mentioned in this section, namely the exponential and the Pareto distribution, with the benefit of relying on the deeper theoretical roots discussed in Section II.

However, giving up theoretical arguments in favor of experimental performances, whatever the approach, the ultimate decision on model is commonly demanded by the authors to various goodness-of-fit (GoF) tests. Aiming to show the benefits of our model, we both repeat part of these tests and make specific theoretical considerations as a further preference argument. We use the same statistics as in [60] and [64], i.e., the Cramer-von-Mises (CvM) test and the Akaike criterion, respectively. In very essential terms, CvM test bases the acceptance of a given distribution law as the source of the observed data (the null hypothesis) on the mean square differ-

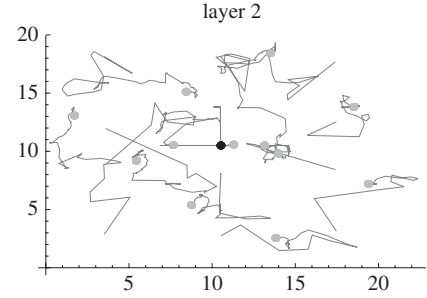


Fig. 13. Mobility tracks of successfully trained neurons. Curve coordinates cartesian coordinates within one hidden layer, bullets: final positions.

ence between the data ECDF and the hypothesized distribution CDF [65]. By contrast the Akaike criterion is a tool to compare the acceptability of various distributions in the above role, it consists of a balanced sum of the data likelihood, as a witness of the questioned distribution fitness, and a second term penalizing the distribution complexity [66]. Thus in Table III, we index the rows with the experimental benchmarks considered in this paper and head the columns with the candidate distributions. In each cell we report: 1) the fraction of traces not rejected by the CvM test having the column distribution as null hypothesis, and 2) the fraction of traces whose Akaike statistic computed on the column distribution proves to be the best (numerically, the lowest in a reverse scale) re the other candidates. To be compliant with analogous tables in the cited papers, we used the maximum likelihood estimators of the competitor distribution laws. As for the latter, we mention that their computation deserves some numerical instability when referred to the tapered distribution [58].

We remark the general superiority of our distribution, which is partly due to the fact that for each trace we have concretely available a huge set of *compatible* distributions from which the one that minimizes the test statistic can be selected. While the Akaike criterion is always in our favor, the CvM test promotes the truncated Pareto distribution on the HMTD benchmark and the exponential distribution on the PTR benchmark. Note that on the other two benchmarks, we also beat the other candidates on the basis of the median curves.

The difference between the two test targets mainly highlights the distinguishing features in favor of our approach versus most alternative ones. Actually, the various artifices to embed an elbow in the Pareto distribution, such as by

multiplying it with an exponential distribution (the origin of tapered distribution), or truncating the former to its maximum observed value (the origin of the truncated Pareto), have the main effect of canceling the heavy tailed feature. As a consequence they miss the fitting of the tail observations, which nevertheless constitute almost one third of the entire sample. Thus, losing the CvM test in the mentioned benchmarks, despite winning w.r.t. the Akaike criterion, may denote the unsuitability of this test for our inference goal rather than a failure of the inference. In any case, since these statistics are distribution free, we may state that, over a sample of 288 elements, the GoF a general walker mobility through our model passes the CvM test in 81.52% of the cases and beats the competitor models on the basis of Akaike statistic 100% of the time.

Summarizing this comparison with state-of-the-art models, we can say that ours gives rise to an intercontact distribution law which both has a strong rationale and meets different fitting criteria. We often beat competitor models as for standard GoF tests, with the additional ability to earn comparable scores even under the unprecedented constraint of preserving the heavy tailed feature of the observed data.

### VIII. CONCLUSION

What was better than a neural network? A network of mobile neurons. With this lead idea in mind, in this paper we present some timing aspects concerning the additional feature of this new kind of connectionist paradigm. We started by highlighting the relevance of these aspects in various (biological, social, and simulation) concerns of the paradigm. Then we explored the modeling matter *per se*, first looking for both theoretical and experimental validation of the wait-and-chase model, and subsequently contrasting our results with other stochastic processes' devised to describe similar operational fields. Herein we claim the original contribution of our work, in terms both of the new connectionist paradigm and of sound technical results in modeling and identification of point processes with memory.

However, we have an efficient neural network training procedure as the medium-term goal of our work, and the premising framework of our paper as well. Thus we conclude this paper with the screenshot shown in Fig. 13, along with some related considerations aimed at indicating a principled applicability of the above results to the asserted framework. The figure reports the neuron mobility paths in a 2-D Euclidean plane. They refer to the second out of a four-layer neural network trained to learn the Pumadyn 8-nm benchmark [67]. Omitting details on the neuron dynamics - a matter requiring in any case an in-depth analysis - we simply remark that the trajectories of the single neurons were generally characterized by an initial random wandering plus a path heavily biased toward a set point which they approach with very contained oscillations. In these traits we may recognize the wait-and-chase phases of our model. As further details, we mention that during these trajectories the neurons had the approach toward the most informative (back-propagating the highest error terms) neurons of the

upper layer as their target, albeit disturbed by the repulsive strength of similar neurons lying on the same layer. As for the temporal dynamics, we have found that the reckoning of the corresponding rendezvous times between neurons of contiguous levels was well synthesized by the enunciated shifted-Pareto distribution law. In regard to key features of our statistical framework, we recognize the non independence of the noise effects along the training process. We also used the inversion methods of Section VI for the initialization of the mobility parameters, whose refinement was carried out as a side target of the overall training procedure. These were nevertheless very preliminary, mainly prospective, considerations, whereas a good assessment of the learning rules in conjunction with a proper setting of the neural mobility timing remains an open problem to be solved in a future work.

### REFERENCES

- [1] O. Marín and G. Lopez-Bendito, "Neuronal migration," in *Evolution of Nervous Systems, Four-Volume Set: A Comprehensive Reference*, J. H. Kaas, Ed., 1st ed. New York: Academic, 2007, ch. 1.1.
- [2] G. A. Carpenter and S. Grossberg, "Adaptive resonance theory," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., 2nd ed. Cambridge, MA: MIT Press, 2003, pp. 87–90.
- [3] M. Bortman and M. Aladjem, "A growing and pruning method for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 20, no. 6, pp. 1039–1045, Jun. 2009.
- [4] A. R. Kriegstein and S. C. Noctor, "Patterns of neuronal migration in the embryonic cortex," *Trends Neurosci.*, vol. 27, no. 7, pp. 392–399, Jul. 2004.
- [5] J. G. Gleeson and C. A. Walsh, "Neuronal migration disorders: From genetic diseases to developmental mechanisms," *Trends Neurosci.*, vol. 23, no. 8, pp. 352–359, 2000.
- [6] O. Marín and J. L. Rubenstein, "Cell migration in the forebrain," *Rev. Neurosci.*, vol. 26, pp. 441–483, Mar. 2003.
- [7] D. Schuler, "Social computing," *Commun. ACM*, vol. 37, no. 1, pp. 28–29, 1994.
- [8] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [9] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *J. Mach. Learn. Res.*, vol. 10, pp. 1–40, Jan. 2009.
- [10] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1. Cambridge, MA: MIT Press, 1986.
- [11] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes (Probability and Its Applications)*, vol. 2, 2nd ed. New York: Springer-Verlag, 2008.
- [12] H. Mori, "Transport, collective motion, and Brownian motion," *Progr. Theoret. Phys.*, vol. 33, no. 3, pp. 423–455, 1965.
- [13] M. Kac, "Random walk and the theory of Brownian motion," *Amer. Math. Monthly*, vol. 54, no. 7, pp. 369–391, Aug.–Sep. 1947.
- [14] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Reading, MA: Addison-Wesley, 1968.
- [15] M. F. Shlesinger, J. Klafter, and G. Zumofen, "Above, below and beyond Brownian motion," *Amer. J. Phys.*, vol. 67, no. 12, pp. 1253–1259, 1999.
- [16] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, Jun. 2008.
- [17] A. M. Edwards, R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. E. Luz, E. P. Raposo, H. E. Stanley, and G. M. Viswanathan, "Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer," *Nature*, vol. 449, pp. 1044–1049, Oct. 2007.
- [18] A. S. Chaves, "A fractional diffusion equation to describe Lévy flights," *Phys. Lett. A*, vol. 239, nos. 1–2, pp. 13–16, 1998.
- [19] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics (Probability and Mathematics Statistics)*. New York: Wiley, 1976.

- [20] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemp. Phys.*, vol. 46, no. 5, pp. 323–351, Sep. 2005.
- [21] R. J. C. Brown, "The collision rate in a dilute classical gas," *Canad. J. Chem.*, vol. 44, no. 12, pp. 1421–1426, 1966.
- [22] F. Bartumeus, F. Peters, S. Pueyo, C. Marrasé, and J. Catalan, "Helical Lévy walks: Adjusting searching statistics to resource availability in microzooplankton," *Proc. Nat. Acad. Sci.*, vol. 100, no. 22, pp. 12771–12775, Oct. 2003.
- [23] A. Chisholm and M. Tessier-Lavigne, "Conservation and divergence of axon guidance mechanisms," *Current Opinion Neurobiol.*, vol. 9, no. 5, pp. 603–615, Oct. 1999.
- [24] G.-L. Ming, S. T. Wong, J. Henley, X.-B. Yuan, H.-J. Song, N. C. Spitzer, and M.-M. Poo, "Adaptation in the chemotactic guidance of nerve growth cones," *Nature*, vol. 417, pp. 411–418, May 2002.
- [25] H. Supér, J. A. Del Río, A. Martínez, P. Pérez-Sust, and E. Soriano, "Disruption of neuronal migration and radial glia in the developing cerebral cortex following ablation of Cajal-Retzius cells," *Cereb. Cortex*, vol. 10, no. 6, pp. 602–613, 2000.
- [26] F. Mann, W. A. Harris, and C. E. Holt, "New views on retinal axon development: A navigation guide," *Int. J. Develop. Biol.*, vol. 48, nos. 8–9, pp. 957–964, 2004.
- [27] L. F. Agnati, D. Guidolin, M. Guescini, S. Genedani, and F. Kjell, "Understanding wiring and volume transmission," *Brain Res. Rev.*, vol. 64, no. 1, pp. 137–159, Sep. 2010.
- [28] J. Lerner, "Role assignments," in *Network Analysis, Methodological Foundations* (Lecture Notes in Computer Science), vol. 3418. Berlin, Germany: Springer-Verlag, 2005, ch. 9, pp. 216–252.
- [29] K. M. Carley, J. Diesner, J. Reminga, and M. Tsvetov, "Toward an interoperable dynamic network analysis toolkit," *Decis. Support Syst.*, vol. 43, no. 4, pp. 1324–1347, Aug. 2007.
- [30] K. T. Ozan and G. Ferrari, *Ad Hoc Wireless Networks: A Communication-Theoretic Perspective*. New York: Wiley, 2006.
- [31] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology* (Texts in Applied Mathematics), vol. 40. New York: Springer-Verlag, 2001.
- [32] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research," in *Proc. REALMAN 2nd Int. Workshop Multi-Hop Netw.*, 2006, pp. 31–38.
- [33] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Comput. Netw.*, vol. 33, nos. 1–6, pp. 309–320, Jun. 2000.
- [34] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, Jan. 2002.
- [35] G. U. Yule, "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.," *Phil. Trans. Royal Soc. London, Ser. B*, vol. 213, nos. 402–410, pp. 21–87, Jan. 1925.
- [36] W. J. Reed, "The Pareto law of incomes—an explanation and an extension," *Phys. A: Stat. Mech. Appl.*, vol. 319, pp. 469–486, Mar. 2003.
- [37] B. B. Mandelbrot, "An information theory of the statistical structure of language," in *Proc. Symp. Appl. Commun. Theory*, London, U.K., Sep. 1952, pp. 486–500.
- [38] V. F. Pisarenko and D. Sornette, "Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution," *Pure Appl. Geophys.*, vol. 160, no. 12, pp. 2343–2364, 2003.
- [39] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.
- [40] W. J. Reed, "The Pareto, Zipf and other power laws," *Econ. Lett.*, vol. 74, no. 1, pp. 15–19, Dec. 2001.
- [41] J. Pickands, "Statistical inference using extreme order statistics," *Ann. Stat.*, vol. 3, no. 1, pp. 119–131, Jan. 1975.
- [42] V. Pareto, *Course d'Economie Politique*. Lausanne, Switzerland: Rouge, 1897.
- [43] P. Arena, L. Fortuna, M. Frasca, and L. Patané, "Learning anticipation via spiking networks: Application to navigation control," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 202–216, Feb. 2009.
- [44] T. Kohonen, *Self-Organizing Maps* (Information Science), vol. 30, 3rd ed. New York: Springer-Verlag, 2001.
- [45] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and the consequence of human mobility in conference environments," in *Proc. ACM WDTM*, 2005, pp. 244–251.
- [46] B. Apolloni, D. Malchiodi, and S. Gaito, *Algorithmic Inference in Machine Learning*, 2nd ed. Magill, Australia: AKI, 2006.
- [47] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Freeman, NY: Chapman & Hall, 1993.
- [48] B. Apolloni, A. Esposito, D. Malchiodi, C. Orovas, G. Palmas, and J. Taylor, "A general framework for learning rules from data," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1333–1349, Nov. 2004.
- [49] B. Apolloni, S. Bassis, S. Gaito, and D. Malchiodi, "Bootstrapping complex functions," *Nonlin. Anal.: Hyb. Syst.*, vol. 2, no. 2, pp. 648–664, Jun. 2008.
- [50] B. Apolloni, S. Bassis, S. Gaito, and D. Malchiodi, "Appreciation of medical treatments by learning underlying functions with good confidence," *Current Pharm. Design*, vol. 13, no. 15, pp. 1545–1570, 2007.
- [51] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.
- [52] J. Lamperti, *Stochastic Processes: A Survey of the Mathematics Theory* (Applied Mathematics Sciences), vol. 23. New York: Springer-Verlag, 1977.
- [53] P. Hall and A. H. Welsh, "Limit theorems for the median deviation," *Ann. Inst. Stat. Math.*, vol. 37, no. 1, pp. 27–36, 1985.
- [54] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.
- [55] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskove, "Mobile call graphs: Beyond power-law and lognormal distributions," in *Proc. 14th ACM SIGKDD*, 2008, pp. 596–604.
- [56] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 606–620, Jun. 2007.
- [57] M. Mitzenmacher, "Dynamic models for file sizes and double Pareto distributions," *Internet Math.*, vol. 1, no. 3, pp. 305–333, 2004.
- [58] Y. Y. Kagan and F. Schoenberg, "Estimation of the upper cutoff parameter for the tapered Pareto distribution," *J. Appl. Probab.*, vol. 38A, pp. 158–175, 2001.
- [59] H. Cai and D. Y. Eun, "Crossing over the bounded domain: From exponential to power-law inter-meeting time in MANET," in *Proc. 13th Annu. Int. Conf. Mobile Comput. Netw.*, Montreal, QC, Canada, 2007, pp. 1578–1591.
- [60] S. Hong, I. Rhee, S. J. Kim, K. Lee, and S. Chong, "Routing performance analysis of human-driven delay tolerant networks using the truncated Lévy walk model," in *Proc. 1st ACM SIGMOBILE Workshop Mobility Models*, 2008, pp. 25–32.
- [61] J.-Y. Le Boudec and M. Vojnović, "Perfect simulation and stationarity of a class of mobility models," in *Proc. IEEE 24th Annu. Joint Conf. Comput. Commun. Soc.*, vol. 4, Mar. 2005, pp. 2743–2754.
- [62] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the Lévy-walk nature of human mobility," in *Proc. INFOCOM 27th Conf. Comput. Commun.*, Phoenix, AZ, Apr. 2008, pp. 924–932.
- [63] M. Piórkowski, "Sampling urban mobility through on-line repositories of GPS tracks," in *Proc. 1st ACM Int. Workshop HotPlanet*, 2009, pp. 1–6.
- [64] V. Conan, J. Leguay, and T. Friedman, "Characterizing pairwise inter-contact patterns in delay tolerant networks," in *Autonomics*. Brussels, Belgium: ICST, 2007, pp. 1–9.
- [65] D. A. Darling, "The Kolmogorov-Smirnov, Cramer-von Mises tests," *Ann. Math. Stat.*, vol. 28, no. 4, pp. 823–838, Dec. 1957.
- [66] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
- [67] P. I. Corke, "A robotics toolbox for MATLAB," *IEEE Robot. Autom. Mag.*, vol. 3, no. 1, pp. 24–32, Mar. 1996.



**Bruno Apolloni** is a Full Professor in computer science with the University of Milan, Milan, Italy. His current research interests include the frontier area between probability and mathematical statistics and computer science with special interest in pattern recognition and multivariate data analysis, probabilistic analysis of algorithms, subsymbolic and symbolic learning processes, and fuzzy systems.





**Simone Bassis** is an Assistant Professor with the Department of Computer Science, University of Milan, Milan, Italy. His current research interests include the inference of spatial and temporal processes, including linear and nonlinear statistical regression, fractal processes identification, dimensionality reduction, multivariate dependence modeling, and evolutionary dynamics applied to both natural and artificial systems.



**Gian Paolo Rossi** is a Full Professor of computer networks with the Computer Science Department, University of Milan, Milan, Italy, where he leads the Computer Networks Laboratory. His current research interests include network communication include architecture and protocol designs, performance evaluation, modeling and analysis, delay/disruption tolerant networks, vehicle-to-vehicle and sensor networks, and on semantic query routing.



**Elena Pagani** is an Associate Professor with the Computer Science Department, University of Milan, Milan, Italy, since 2006. Her current research interests include network protocols and architectures, wireless technologies, and performance evaluation.



**Lorenzo Valerio** is currently pursuing the Ph.D. degree with the Department of Mathematics, University of Milan, Milan, Italy. His current research interests include the inference of spatial and temporal processes, the connection between mobility, and computational learning models, and finally mobile technologies such as mobile networks and mobile operating systems.