Hidden scaling patterns and universality in written communication

M. Formentin,[1,*] A. Lovison,[2,†] A. Maritan,[1,‡] and G. Zanzotto[3,§]
[1]*Dipartimento di Fisica "Galileo Galilei", Università degli studi di Padova, Via Marzolo 8, 35131 Padova, Italy*
[2]*Dipartimento di Matematica, Università di Padova, via Trieste 63, I-35121 Padova, Italy*
[3]*Dipartimento di Psicologia Generale, Università di Padova, via Venezia 12, I-35131 Padova, Italy*
(Received 20 February 2014; published 31 July 2014)

The temporal statistics exhibited by written correspondence appear to be media dependent, with features which have so far proven difficult to characterize. We explain the origin of these difficulties by disentangling the role of spontaneous activity from decision-based prioritizing processes in human dynamics, clocking all waiting times through each agent's "proper time" measured by activity. This unveils the same fundamental patterns in written communication across all media (letters, email, sms), with response times displaying truncated power-law behavior and average exponents near $-\frac{3}{2}$. When standard time is used, the response time probabilities are theoretically predicted to exhibit a bimodal character, which is empirically borne out by our newly collected years-long data on email. These perspectives on the temporal dynamics of human correspondence should aid in the analysis of interaction phenomena in general, including resource management, optimal pricing and routing, information sharing, and emergency handling.

DOI: 10.1103/PhysRevE.90.012817 PACS number(s): 89.75.Da, 02.50.Ga, 02.50.Le, 89.65.Ef

I. INTRODUCTION

Remarkable statistical regularities observed in human and animal dynamics have attracted much attention in recent years [1–14]. A particularly interesting and studied case is given by written communication, which, whether on paper ("letters") or in electronic form ("email"), is a most fundamental human activity [10,15–23], sustaining and giving the tempo to much of our civilization's advance. In recent times short-text messaging ("sms") has also been added to the repertoire of media through which humans intensely communicate with each other in writing [24].

A main feature of interactive processes such as written correspondence is that, regardless of medium, the behavior and temporal dynamics of any agent $\mathcal{A}$ are characterized by two distinct waiting times, i.e., response times (RTs) and interevent times (IETs), schematically represented in Fig. 1; see also the Supplemental Material (SM) for precise definitions [25]. We denote the probability distributions of RTs and IETs respectively by $P_R(\tau)$ and $P_I(\tau)$, where $\tau = \Delta t \in \mathbb{N}^+$ is the length of time intervals (with time t measured in days for letters, and seconds for email and sms). A better understanding of the mechanisms at the basis of written communication thus entails the analysis of these waiting times within large-scale interaction networks whose overall dynamics is largely unknown. During the last decade these and related questions have attracted the attention of a research community going from mathematics to physics to sociology, whose studies, grounded on a number of databases which collect basic empirical information on communication events, have begun to clarify some basic facts on the behavior of such networks and the agents in it. In the SM we give details about the communication datasets used for the present work (denoted DL1, DE1, etc.; see Table I), which include data previously available on written correspondence (letters, email, sms), as well as two new long-term email datasets collected for the present study.

II. STATE OF THE ART ON TIME DISTRIBUTIONS AND CONTROVERSY

The first notable observation derived from the analysis of the empirical data is that events for all communication media occur in a highly intermittent fashion, with time fluctuations producing heavy-tailed distributions for both $P_I(\tau)$ and $P_R(\tau)$. The characterization of these statistics has been strongly debated, as they appear to depend on the medium (letters, email, sms) and lack universal features [10,15,17–20,22–24,26,27], although the investigation in Ref. [19] led to a form of universality for the IETs in letters and emails. In spite of earlier indications of scaling for the empirical distributions $P_R(\tau)$ with two different exponents, -1 and $-\frac{3}{2}$, respectively in email and letters [10,15,18,20,22,26], the scaling nature and general features of $P_R(\tau)$ for email are still contrastingly judged [27]. Different priority queueing models have also been used to account for these controversial observations, producing power-law behavior for $P_R(\tau)$ with theoretical exponents -1 or $-\frac{3}{2}$ (see Refs. [10,15,28–30]), as well as exponents varying in a range from -1 to under -2 (Refs. [15,31–40]).

III. RECLOCKING THE PROBABILITY DISTRIBUTIONS THROUGH ACTIVITY

To shed light on these poorly understood aspects of written communication, we disentangle from the overall time dynamics of a given agent $\mathcal{A}$ the contributions due to $\mathcal{A}$'s spontaneous interevent pauses. To do this we introduce the parameter $s \in \mathbb{N}^+$ which counts the number of $\mathcal{A}$'s outgoing communication events (a measure of $\mathcal{A}$'s activity), so that each increase by one unit for s corresponds to an IET for $\mathcal{A}$; see Fig. 1. The probability densities for both the RTs and IETs, which characterize $\mathcal{A}$'s behavior, can be computed in

*marco.formentin@rub.de
†lovison@math.unipd.it
‡maritan@pd.infn.it
§giovanni.zanzotto@unipd.it

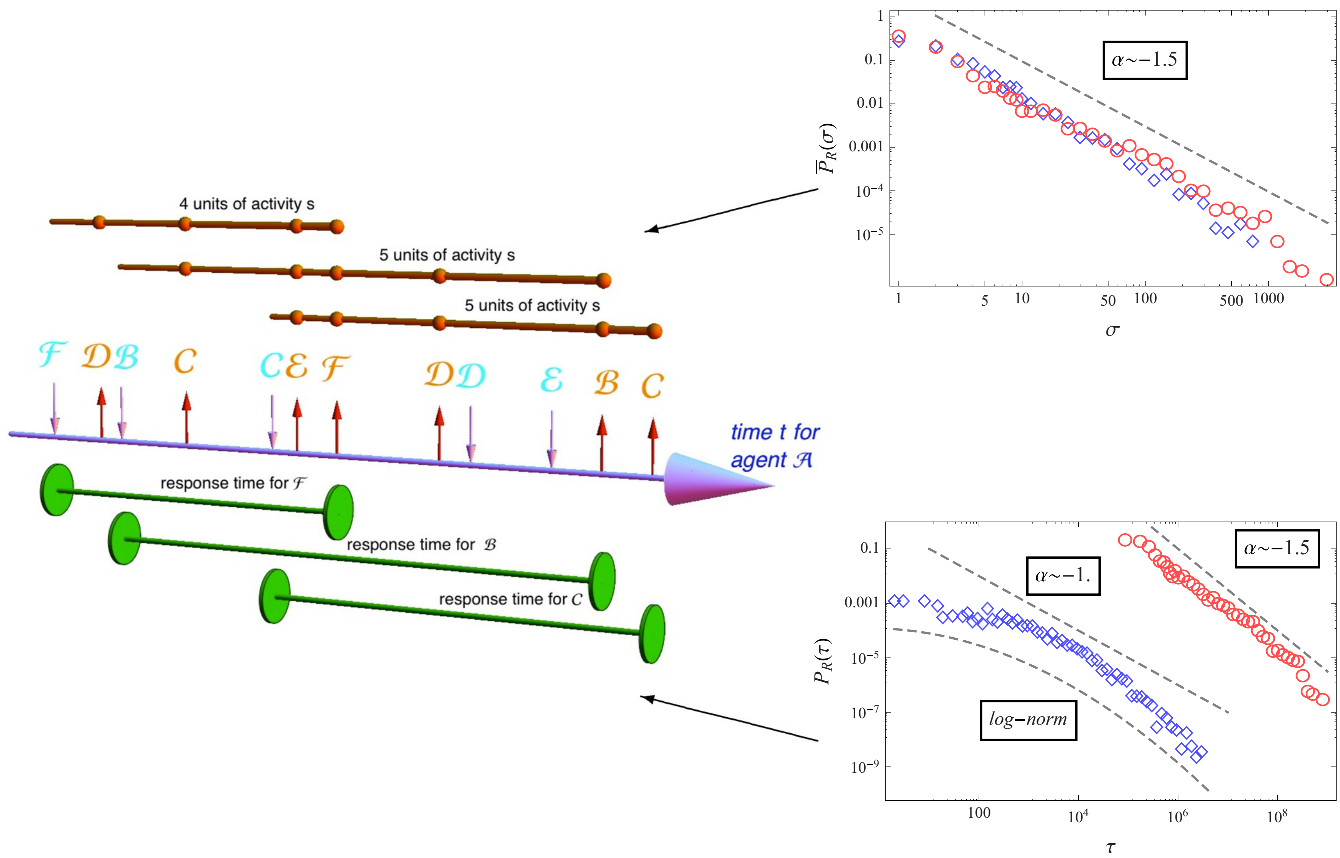

FIG. 1. (Color online) Two clocks for written correspondence. Representation of the communication activity along the axis of time t for an agent $\mathcal{A}$. Arrows pointing into the t axis mark incoming messages from the indicated agents $\mathcal{B}$, $\mathcal{C}$, etc., arrows pointing out of the t axis mark response messages to the same agents. The intervals between such arrows define the interevent times (IETs) of agent $\mathcal{A}$. The response times (RTs) of $\mathcal{A}$ are defined as shown, either clocked through time t (all measured in seconds), or through the activity parameter s which counts the number of outgoing messages from $\mathcal{A}$ (see also the Supplemental Material). The associated RT probability distributions are denoted by $P_R(\tau)$ and $\bar{P}_R(\sigma)$ when clocked respectively through s or t (with $\tau = \Delta t$ and $\sigma = \Delta s$). The RT distributions in terms of t are nonuniversal, as they depend on the communication medium and the agent; see the lower diagram for $P_R(\tau)$, showing the t-clocked RTs of representative agents communicating through letters (red circles) and email (blue diamonds). In contrast, we find that the same RTs, when clocked through activity s, give distributions as in the upper diagram for $\bar{P}_R(\sigma)$, which are almost superposable power laws following Eq. (4.1), with individual exponents α on average near $-\frac{3}{2}$ for all media (letters, email, sms).

terms of $\sigma = \Delta s$ in place of $\tau = \Delta t$. In analogy to similar clocking alternatives arising for instance in special relativity, the parameter s can be interpreted, up to a suitable scale factor, as an agent's "proper time"; the introduction of s bears also a relation to the "events per active interval" considered for different purposes in Ref. [19]. We denote by $\bar{P}_R(\sigma)$ and $\bar{P}_I(\sigma)$ the s-clocked probability distributions for the RTs and IETs respectively, and notice that the s-clocked IET distribution $\bar{P}_I(\sigma)$ is trivially the same for all agents and media, being concentrated by definition at $\sigma = 1$. See Eq. (5.1) below and the SM for details on the mathematical relation among the probabilities $P_R(\tau)$, $P_I(\tau)$, and $\bar{P}_R(\sigma)$.

IV. POWER-LAW EMPIRICAL PROBABILITIES AFTER RECLOCKING

Remarkably, we find that in all databases, across all media, the RTs of active agents, when clocked through activity s, are described by discrete exponentially truncated power laws of the form

$$\bar{P}_R(\sigma) \sim \sigma^{\alpha} e^{-\sigma/\lambda}, \qquad (4.1)$$

where α is the scaling exponent, and λ is the cutoff parameter [41]. A number of empirical distributions $\bar{P}_R(\sigma)$ as in Eq. (4.1), representative of the s-clocked RTs for each written communication medium (letters, email, sms), are shown in Fig. 2 (see the SM for more statistics). The individual exponents in the empirical RT distributions in Eq. (4.1) have average values close to $-\frac{3}{2}$ for all three media, as detailed in Table I (the thresholds $\sigma_{\min}$ for the fitting are given in Table II). The truncated scaling in Eq. (4.1) of $\bar{P}_R(\sigma)$ with exponents averaging near $-\frac{3}{2}$ can be clearly appreciated also in the most active sms agents, despite their having comparatively much scarcer statistics than in email or letters. For email, these results on the scaling of $\bar{P}_R(\sigma)$ and its exponents are validated in agents across all three independently collected databases. For confirmation, we also sampled the long-term email data through 3-, 6-, 12-, and 18-month windows within the total two-year period of dataset DE1. The RT distributions in the

TABLE I. On the left are indicated the databases analyzed in this work for the three written-communication media (letters, email, sms); see the SM for details. On the right are reported the corresponding exponents α computed for the empirical RT probabilities $\bar{P}_R(\sigma)$ in Eq. (4.1), clocked through activity s. Individual values of α are given for databases DL1, DE2; the average $\bar{\alpha}$ and standard deviation σ of the distributions of individual exponents are indicated for the databases DE1, DE3, DS1.

Medium	Database	Exponents
Letters	DL1: agents CD, AE, SF	$\alpha_{CD} = 1.493 \pm 0.020$ $\alpha_{AE} = 1.565 \pm 0.013$ $\alpha_{SF} = 1.886 \pm 0.028$
Email	DE1: new two-year database	$\bar{\alpha} = 1.543$ $\sigma = 0.306$
	DE2: new very long term database, agents AL, AP, FC	$\alpha_{AL} = 1.539 \pm 0.024$ $\alpha_{AP} = 1.557 \pm 0.011$ $\alpha_{FC} = 1.478 \pm 0.008$
	DE3: three-month database from Ref. [18]	$\bar{\alpha} = 1.562$ $\sigma = 0.366$
sms	DS1: one-month database from Ref. [24]	$\bar{\alpha} = 1.447$ $\sigma = 0.444$

three-month database DE3 resulted to be superposable with those obtained from the three-month sampling of the long-term databases DE1 and DE2; furthermore, the average individual exponents obtained in this way were found to be near $-\frac{3}{2}$ for all window lengths, within and across the three email datasets [see the SM, Figs. 6 and 7(b)].

Summarizing, while the waiting time distributions may vary across agents and media when expressed in terms of standard time t, all waiting times have quite the same medium-independent form when computed through proper time s, with a definite convergence of the exponents to average values near $-\frac{3}{2}$ in all media. This goes together with the (trivial) universality of the s-clocked IET distributions $\bar{P}_I(\sigma)$, which are all concentrated at $\sigma = 1$ as mentioned earlier. The introduction of the activity clocking thus emphasizes an intrinsic universal component underlying all written communication, partly obfuscated by the interaction with the spontaneous

TABLE II. Numerical simulations vs empirical data: p values obtained from KS tests for datasets DE1 (email, two years), DE2 (email, long term), DL1 (letters).

Database	$\sigma_{\min}$	$p > 0.05$	$p > 0.01$
DE1	10	76.6%	85.6%

Database	$\sigma_{\min}$	p
DE2		
AL	2	0.62
AP	3	0.61
FC	3	0.55
DL1		
CD	2	0.50
AE	20	0.67
SF	3	0.92

IETs, which are media- and agent-dependent. We discuss such universality more in detail below.

V. BIMODAL EMPIRICAL PROBABILITIES CLOCKED THROUGH TIME

In the light of the above results on the s-clocked distributions $\bar{P}_R(\sigma)$, we can now better analyze the empirical t-clocked RT distributions $P_R(\tau)$ of written correspondence. For the same agents as in Fig. 2, and for each medium (letters, email, sms), the individual $P_R(\tau)$ are represented in Fig. 3, the insets showing the associated IET distributions $P_I(\tau)$. We see from Fig. 3 that the t-clocked RT distributions $P_R(\tau)$ do not scale, and exhibit complex, media-dependent characteristics (more statistics are reported in the SM).

This behavior of $P_R(\tau)$ can be understood by considering that the t-clocked distribution $P_R(\tau)$ of any agent $\mathcal{A}$ can be retrieved in a natural way by compounding the IET probabilities $P_I(\tau)$, characterizing the spontaneous action of $\mathcal{A}$, back into the s-clocked RT power law $\bar{P}_R(\sigma)$ in Eq. (4.1), i.e., by separating any two consecutive activities of $\mathcal{A}$ through random time intervals sampled from the IET distribution $P_I(\tau)$ of $\mathcal{A}$ (representative examples of IET distributions in the different media are shown in the insets of Fig. 3). Specifically, let $N \sim \bar{P}_R(\sigma)$ and $\rho_I(h) \sim P_I(\tau)$, $h = 1,2,\ldots,$ be independent random variables, with N giving the number of activities between a message reception by $\mathcal{A}$ and the response to it; then, the t-clocked RTs for $\mathcal{A}$ are described by the compounding process

$$\rho_R = \sum_{h=1}^{N} \rho_I(h) \quad \text{with law}$$
$$P_R(\tau) = \sum_{\sigma \geqslant 1} \text{Prob}\left(\sum_{h=1}^{\sigma} \rho_I(h) = \tau\right) \bar{P}_R(\sigma). \qquad (5.1)$$

We have checked through numerical simulations that the above relation holds for the empirical distributions $P_R(\tau)$, $P_I(\tau)$, $\bar{P}_R(\sigma)$. This agreement between the empirical data and Eq. (5.1), in which the IETs are assumed to be independent random variables, indicates implicitly that correlations in the waiting times of human correspondence, if any, do not significantly affect the compounding of probabilities in Eq. (5.1). This also agrees with the results in Ref. [43] indicating a lack of correlations within the IET statistics from the email data in Ref. [18].

In the SM we show that the t-clocked RT distributions $P_R(\tau)$ in Eq. (5.1) result to have a *bimodal* character when the IET distribution $P_I(\tau)$ is heavy tailed and $\bar{P}_R(\sigma)$ is scaling as in Eq. (4.1). This can be understood for instance by computing the generating function [44] of the random variable ρ_R in Eq. (5.1), defined as

$$G_{\rho_R}(z) = \sum_{\tau \geqslant 1} P_R(\rho_R = \tau) z^{\tau}, \quad z \in [0,1], \qquad (5.2)$$

which encodes the law of ρ_R, as $P_R(\rho_R = \tau) = \frac{1}{\tau!}\frac{d^{\rho} G_{\tau_R}}{dz^{\tau}}(0)$. For the present purposes, we can assume the simplified forms for the distributions $\bar{P}_R(\sigma) \sim \sigma^{\alpha}$, with α near $-\frac{3}{2}$, and

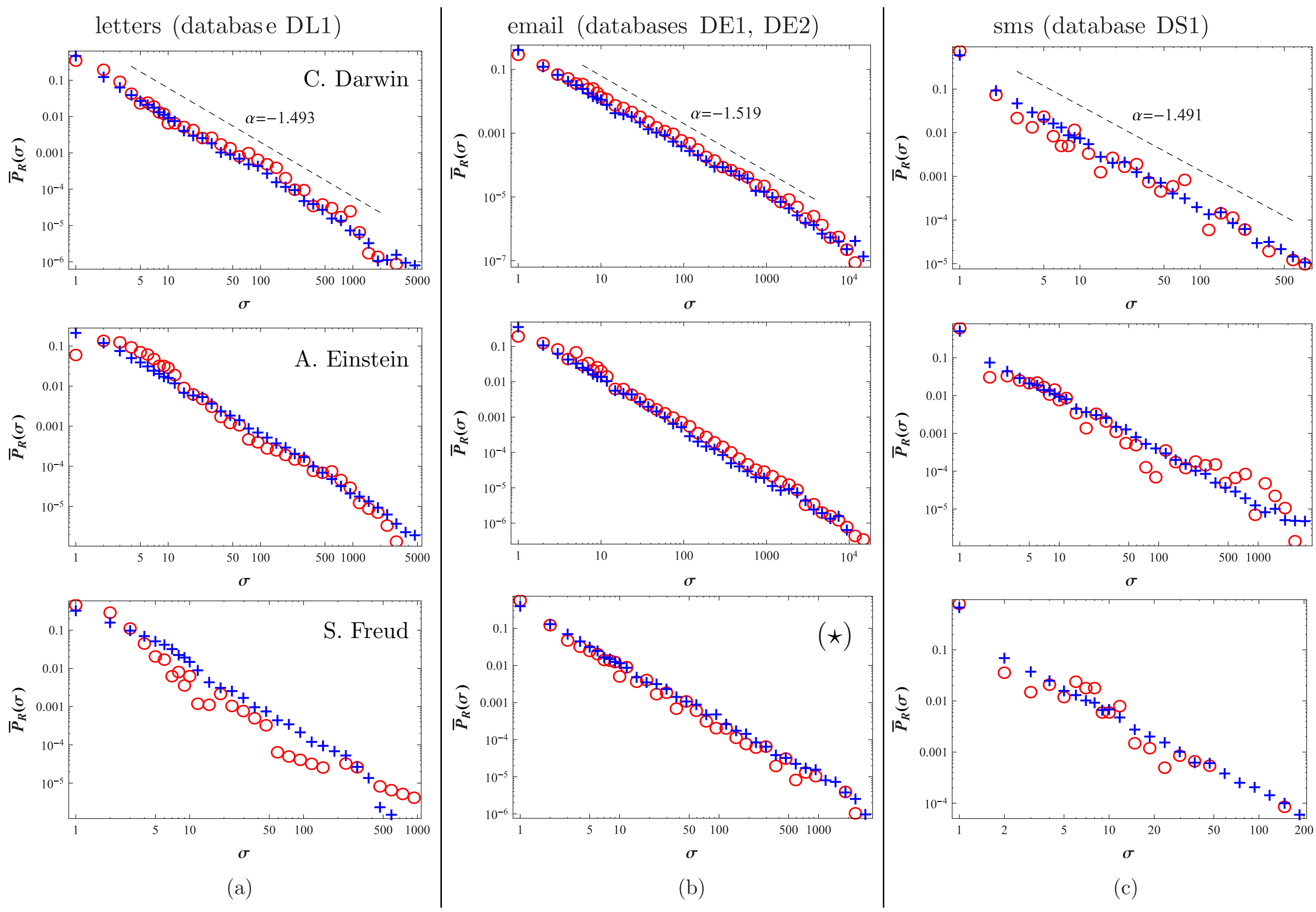

FIG. 2. (Color online) Response times re-clocked through activity. Log-log plots of the response-time probability densities $\bar{P}_R(\sigma)$ clocked through activity s, for three typical agents for each different written-communication medium (logarithmic binning [42]). Red circles indicate empirical data; blue crosses represent our model predictions. (a) Letters: data from database DL1, on the correspondence of C. Darwin, A. Einstein, and S. Freud; (b) email: data from typical agents in the long term databases DE1 and DE2 (the agent in DE2, with data spanning seven years, is marked by an asterisk); (c) sms: data from typical agents in the database DS1 of Ref. [24]. The probability densities for all media are very well fitted by the truncated power laws in Eq. (4.1) with individual exponents α as follows (going from top to bottom in each column): 1.493, 1.565, 1.886 (letters); 1.519, 1.604, 1.539 (email); 1.491, 1.215, 1.097 (sms). See Table I for information on the exponents in the various databases, and the SM for more statistics. The straight dashed lines in the top diagrams are drawn to guide the eye, with the indicated exponents.

$P_I(\tau) \sim \tau^{\beta}\exp(-\tau/T_I)$ for some characteristic time T_I. Then, the analysis in Sec. 13 of the SM shows that, due to Eq. (5.1), for large τ, $P_R(\tau)$ has power-law tails with the same exponent near $-\frac{3}{2}$ as $\bar{P}_R(\sigma)$, while, for small τ, $P_R(\tau)$ is affected by the specific features of $P_I(\tau)$, i.e., it scales with exponent -1 for small τ (see the SM, Fig. 5). The crossover in $P_R(\tau)$ occurs for τ of the order of the characteristic time $T_I \sim \frac{\langle\tau^2\rangle}{\langle\tau\rangle}$ of the empirical IET distributions $P_I(\tau)$. In accordance to such prediction, we see in Fig. 3 that the empirical t-clocked RT probabilities about $P_R(\tau)$ do exhibit media-dependence with a complex, bimodal behavior. The latter is particularly evident in the $P_R(\tau)$ distributions derived from the new long-term data on email, which span the largest number of decades in time, from seconds to several years (databases DE1 and DE2). The bimodality of $P_R(\tau)$ likely led to the controversial conclusions earlier reported in the literature about the time statistics in email communication, which our study now contributes to clarify.

VI. MODELING AND UNIVERSAL MECHANISM

To establish a theoretical basis for the above observations on the time patterns of written communication, we show that both the empirically reported s- and t-clocked statistics (Figs. 2 and 3) can be interpreted through priority queueing. We build on previous work about such modeling for human correspondence [10,15,20,28–32,45,46], and demonstrate that we can obtain both the scaling distributions $\bar{P}_R(\sigma)$ in Eq. (4.1), as well as the bimodal distributions $P_R(\tau)$ derived from Eq. (5.1), once the individual IETs and the message arrival times of each agent are suitably accounted for within a universal prioritization framework.

Let $\mathcal{A}$ be an agent with given IET empirical distribution $P_I(\tau)$ (see the examples in Fig. 3), and assume for $\mathcal{A}$ an initial list of L tasks, whose priorities y are independently sampled from the uniform distribution on $[0,1]$ (consistent with the hypothesis that $\mathcal{A}$ is embedded in a complex communication network producing largely independent stimuli to $\mathcal{A}$). At

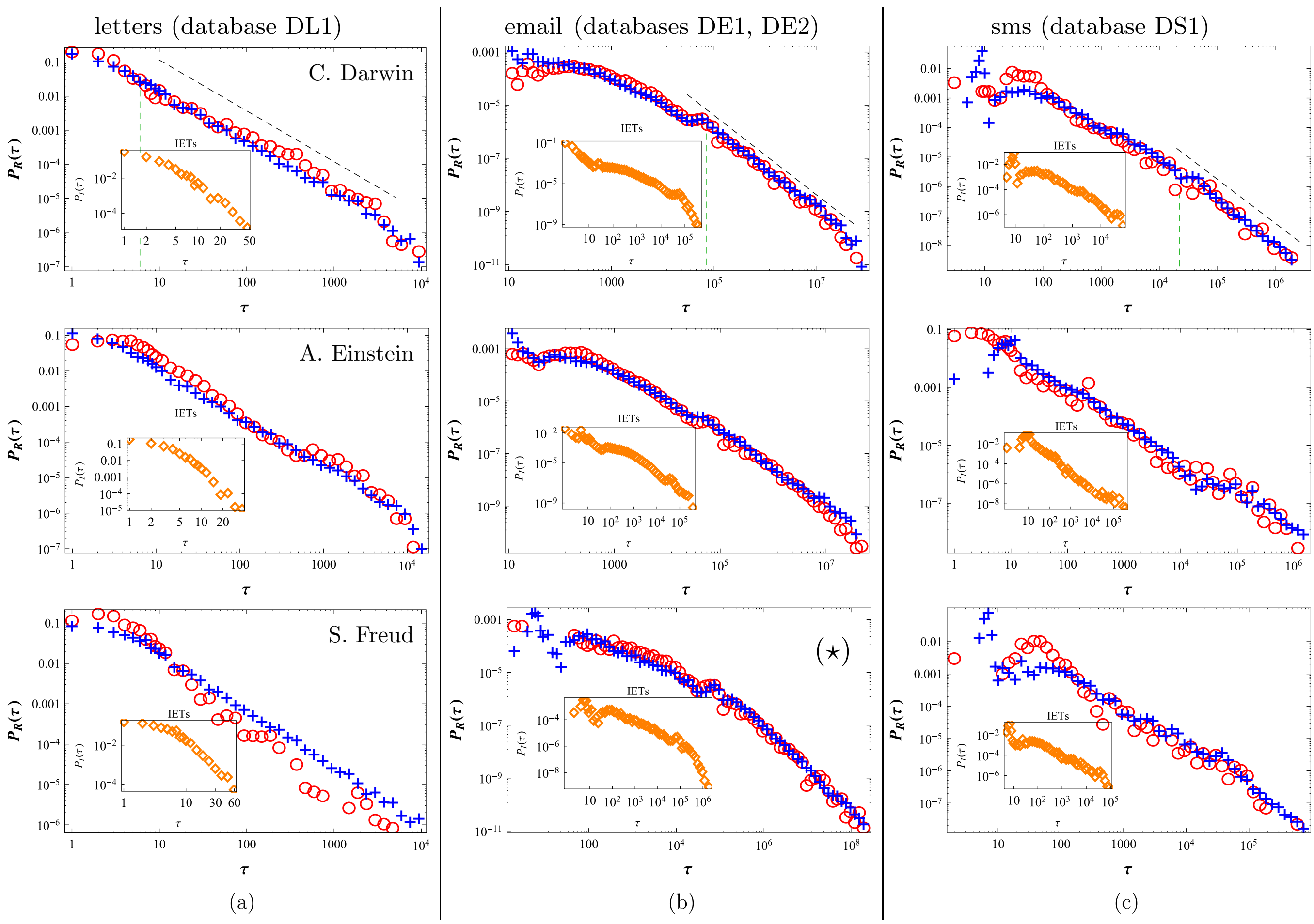

FIG. 3. (Color online) Response times clocked through standard time. Log-log plots of empirical response-time probability densities $P_R(\tau)$ clocked through standard time t (in days for letters, and seconds for email and sms), relative to the same typical agents as in Fig. 2, for all media (logarithmic binning [42]). Red circles indicate empirical data; blue crosses represent computational predictions. See the SM for more statistics. As predicted (see the SM), $P_R(\tau)$ is affected, for small τ, by the specific features of $P_I(\tau)$, while the tails of $P_R(\tau)$ for large τ follow power laws with the same exponents α as the associated s-clocked distributions $\bar{P}_R(\sigma)$, shown in Fig. 2. The bimodality in these t-clocked RT probabilities $P_R(\tau)$ is particularly evident in the RTs for email, in column (b). Also following predictions (see the SM), the crossover in $P_R(\tau)$ occurs for $\tau \sim T_I$ (green dashed vertical lines), where T_I is the characteristic time of the empirical IET distributions $P_I(\tau)$ of each agent, shown in the insets. Typical empirical values are $T_I \sim 10^4$–10^5 sec for email and sms, and $T_I \sim$ 5–10 days for letters.

each time step, corresponding to a unit increment of $\mathcal{A}$'s activity s, the highest priority task in the list is executed (a message replied), and m new tasks are added to the list, each one with priority y sampled as above. The number m is derived at each step by considering the empirical distribution of incoming messages to $\mathcal{A}$ between any two consecutive outgoing messages of $\mathcal{A}$, the data typically giving $m > 1$. The list dynamics in these prioritization process only depends on the task ranking, and is not affected by the hypothesis of a uniform distribution of y values [47]; it was analytically proven that depending on the distribution of m values, the RT distribution $\bar{P}_R(\sigma)$ produced in this way decays for $s \to \infty$ as a power law with exponent at or near $-\frac{3}{2}$ [29,31,32,45,46].

The numerical results for the s-clocked steady-state RT distribution $\bar{P}_R(\sigma)$ for this model are shown in Fig. 2. For all media (letters, email, sms) the simulations follow very closely their empirically observed counterparts, tracing power laws with the correct individual exponents, even when the latter depart considerably from their average value near $-\frac{3}{2}$. Systematic Kolmogorov-Smirnov (KS) tests for discrete distributions [41,48–50] confirm the strong statistical agreement of the simulated s-clocked RTs obtained from prioritization with the corresponding empirical data for each agent, as reported in Table II (see also the SM).

The model also accounts for the bimodality of the t-clocked RT statistics of human correspondence, reported in Fig. 3. The distribution $P_R(\tau)$ of each agent can again be derived from the computed $\bar{P}_R(\sigma)$ as in Fig. 2, by separating, as in Eq. (5.1), the activity events of $\mathcal{A}$ through random time intervals sampled from the empirical IET distribution $P_I(\tau)$ pertaining to $\mathcal{A}$. The log-log plots of the distributions $P_R(\tau)$ so obtained are shown in Fig. 3. We see that, apart from the shortest RTs (10–20 sec) in the electronic media, the numerical predictions are in very good agreement with the empirical results for all media. This confirms that for active agents our approach consistently reproduces very well the empirical data for both the s- and

t-clocked RT distributions across all media in a wide range of exponents averaging near $-\frac{3}{2}$. See also the SM, Figs. 1–4.

VII. CONCLUSIONS

Our findings highlight the interplay between individual spontaneous activity (subsumed by the IET distributions) and universal decision-based processes (subsumed by task prioritization) in the origin of the complex time patterns of written communication. We determine the role of both these factors in the generation of scaling s-clocked RT distributions $\bar{P}_R(\sigma)$ with exponents α near $-\frac{3}{2}$, as well as [through the compounding in Eq. (5.1)] in producing bimodal t-clocked RT distributions $P_R(\tau)$, in very close accordance with empirical data for all media. This gives a different perspective on the nature and features of the temporal inhomogeneities in human dynamics and their underlying mechanisms; in particular, our results explain why earlier views were inadequate regarding the media-dependent power-law or log-norm character of the t-clocked response functions for letters and email, as we bring these two media within the same setting, with text messaging as well.

Interestingly, we see that the power-law behavior in Eq. (4.1) does not arise when written communication occurs mostly in pairs, as analyzed in Ref. [24], because in this case the t-clocked RTs and IETs are strongly correlated, i.e., the s-clocked RTs almost coincide with the s-clocked IETs, being both concentrated near $\sigma = 1$. In contrast, human dynamics with large fluctuations and scaling statistics arises from the operation of complex interaction networks with rich-enough topologies. Prioritization processes then give average values near $-\frac{3}{2}$ to the emerging exponents α, although the latter bear the signature of each agent's input from the network, as the individual deviations of α from $-\frac{3}{2}$ are shown by the model to be affected by the specific arrival-time statistics. To a lesser degree, the exponents may further be influenced by other factors, such as social structure, interest, habit, as discussed in Refs. [33–40,51]. While in our approach the IET distributions $P_I(\tau)$ of agents are derived from the empirical data, various avenues for a theoretical understanding of IETs can be considered, along the lines of Refs. [9,10,17,47,52–54]. The explicit IET fit proposed in Ref. [17] could also be used in Eq. (5.1) to obtain a fully numerical reproduction of the empirical data. This complements our insight into the dynamics of written correspondence as representing the wider network of human interactions, driven by distributed co-operative effects as well as deliberate vs spontaneous individual processes. The proposed methods have wide applicability and may help uncover and analyze hidden patterns also in other contexts where the interaction of human or nonhuman agents alike generates dynamic networks in which discrete (possibly bursty) node activity creates intermittent collective and co-ordinated behaviors.

ACKNOWLEDGMENTS

We thank Dr. J.-P. Eckmann, Dr. M. Gravino, Dr. R. D. Malmgrem, Dr. J. Oliveira, Dr. A. Pellizzon, and three individual long-term email users, for providing us part of the communication data analyzed in this study. A.M. acknowledges the Cariparo Foundation for financial support.

[1] T. Nakamura, T. Takumi, A. Takano, N. Aoyagi, K. Yoshiuchi, Z. R. Struzik, and Y. Yamamoto, PLoS One **3**, e2050 (2008).
[2] K. Hanai, M. Ozaki, D. Yamauchi, and Y. Nakatomi, WSEAS Trans. Biol. Biomed. **3**, 511 (2006).
[3] T. Henderson and S. Bhatti, in *MULTIMEDIA '01: Proceedings of the 9th ACM International Conference on Multimedia* (ACM, New York, NY, 2001), pp. 90–94.
[4] Q. Wang and J.-L. Guo, Physica A **389**, 2127 (2010).
[5] C. Wang and B. A. Huberman, Sci. Rep. **2**, 633 (2012).
[6] Z. Dezso, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A. L. Barabási, Phys. Rev. E **73**, 066132 (2006).
[7] B. Gonçalves and J. J. Ramasco, Phys. Rev. E **78**, 026123 (2008).
[8] L. Gao, J.-L. Guo, C. Fan, and X.-J. Liu, CoRR, 1010–3815.
[9] A. Proekt, J. R. Banavar, A. Maritan, and D. W. Pfaff, Proc. Natl. Acad. Sci. USA **109**, 10564 (2012).
[10] A. Vazquez, J. G. Oliveira, Z. Dezso, K. I. Goh, I. Kondor, and A. L. Barabasi, Phys. Rev. E **73**, 036127 (2006).
[11] A. M. Edwards, R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. E. da Luz, E. P. Raposo, H. E. Stanley *et al.*, Nature (London) **449**, 1044 (2007).
[12] D. Rynski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, Proc. Natl. Acad. Sci. USA **106**, 12640 (2009).
[13] C. Anteneodo and D. R. Chialvo, Chaos **19**, 033123 (2009).
[14] P. Indic, P. Salvatore, C. Maggini, S. Ghidini, G. Ferraro, R. J. Baldessarini, and G. Murray, PLoS ONE **6**, e20650 (2011).
[15] A. L. Barabasi, Nature (London) **435**, 207 (2005).
[16] G. Kossinets and D. J. Watts, Science **311**, 88 (2006).
[17] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. N. Amaral, Proc. Natl. Acad. Sci. USA **105**, 18153 (2008).
[18] J. P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. USA **101**, 14333 (2004).
[19] R. D. Malmgren, D. B. Stouffer, A. S. L. O. Campanharo, and L. A. Nunes Amaral, Science **325**, 1696 (2009).
[20] J. G. Oliveira and A.-L. Barabasi, Nature (London) **437**, 1251 (2005).
[21] Z. D. Zhao, H. Xia, M. S. Shang, and T. Zhou, Chin. Phys. Lett. **28**, 068901 (2011).
[22] S. Qu, Q. Wang, and L. Wang, in *Proceedings of 2011 International Conference on Modelling, Identification and Control (ICMIC)* (IEEE, Shanghai, 2011), pp. 282–284.
[23] D. B. Stouffer, R. D. Malmgren, and L. A. N. Amaral, arXiv:physics/0510216.
[24] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber, Proc. Natl. Acad. Sci. USA **107**, 18803 (2010).
[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.90.012817 for details on databases, definitions, statistical analysis, analytical computations, and additional figures and statistics.
[26] A. Johansen, arXiv:physics/0602029v1.
[27] D. B. Stouffer, R. D. Malmgren, and L. A. N. Amaral, arXiv:physics/0605027v1.

[28] A. Cobham, J. Oper. Res. Soc. Am. **2**, 70 (1954).
[29] J. Abate and W. Whitt, Queueing Syst. **25**, 173 (1996).
[30] G. Grinstein and R. Linsker, Phys. Rev. E **77**, 012101 (2008).
[31] N. Masuda, J. S. Kim, and B. Kahng, Phys. Rev. E **79**, 036106 (2009).
[32] J. Walraevens, T. Demoor, T. Maertens, and H. Bruneel, Phys. Rev. E **85**, 021139 (2012).
[33] P. Blanchard and M.-O. Hongler, Phys. Rev. E **75**, 026102 (2007).
[34] B. Min, K.-I. Goh, and I.-M. Kim, Phys. Rev. E **79**, 056110 (2009).
[35] W.-K. Cho, B. Min, K.-I. Goh, and I.-M. Kim, Phys. Rev. E **81**, 066109 (2010).
[36] K. Kim and K. C. Chae, Eur. J. Oper. Res. **200**, 473 (2010).
[37] R. Crane, F. Schweitzer, and D. Sornette, Phys. Rev. E **81**, 056101 (2010).
[38] T. Maillart, D. Sornette, S. Frei, T. Duebendorfer, and A. Saichev, Phys. Rev. E **83**, 056101 (2011).
[39] A. Saichev and D. Sornette, Phys. Rev. E **81**, 016108 (2010).
[40] H.-H. Jo, R. K. Pan, and K. Kaski, Phys. Rev. E **85**, 066101 (2012).
[41] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Rev. **51**, 661 (2009).
[42] S. Milojević, J. Am. Soc. Inf. Sci. Technol. **61**, 2417 (2010).
[43] C. Anteneodo, R. D. Malmgren, and D. R. Chialvo, Eur. Phys. J. B **75**, 389 (2010).
[44] P. Flajolet and R. Sedgewick, *Analityc Combinatorics* (Cambridge University Press, New York, 2008).
[45] A. Gabrielli and G. Caldarelli, J. Stat. Mech.: Theor. Exp. (2009) P02046.
[46] A. Gabrielli and G. Caldarelli, Phys. Rev. E **79**, 041133 (2009).
[47] A. Vazquez, Phys. Rev. Lett. **95**, 248701 (2005).
[48] B. Rousseau and R. Rousseau, Cybermetrics **4**, 4 (2000).
[49] W. J. Conover, J. Am. Stat. Assoc. **67**, 591 (1972).
[50] T. B. Arnold and J. W. Emerson, The R Journal **3**, 34 (2011).
[51] A. Kentsis, Nature (London) **441**, E5 (2006).
[52] A. Vazquez, Physica A **373**, 747 (2007).
[53] J. G. Oliveira and A. Vazquez, Physica A **388**, 187 (2009).
[54] Y. Jiao, Y. H. Liu, J. Wang, and J. Wang, Chin. Sci. Bull. **55**, 2744 (2010).