



# Communication activity in a social network: relation between long-term correlations and inter-event clustering

Diego Rybski<sup>1,2</sup>, Sergey V. Buldyrev<sup>3</sup>, Shlomo Havlin<sup>4</sup>, Fredrik Liljeros<sup>5,6</sup> & Hernán A. Makse<sup>1</sup>

## SUBJECT AREAS:

STATISTICAL PHYSICS,  
THERMODYNAMICS AND  
NONLINEAR DYNAMICS

STATISTICS

APPLIED PHYSICS

MODELLING AND THEORY

Received

12 April 2012

Accepted

11 July 2012

Published

6 August 2012

Correspondence and  
requests for materials  
should be addressed to  
H.A.M. (hmakse@lev.  
ccny.cuny.edu)

<sup>1</sup>Levich Institute and Physics Department, City College of New York, New York, NY 10031, USA, <sup>2</sup>Potsdam Institute for Climate Impact Research (PIK), P.O. Box 60 12 03, 14412 Potsdam, Germany, <sup>3</sup>Department of Physics, Yeshiva University, New York, NY 10033, USA, <sup>4</sup>Minerva Center and Physics Department, Bar-Ilan University, Ramat Gan 52900, Israel, <sup>5</sup>Department of Sociology, Stockholm University, S-10691 Stockholm, Sweden, <sup>6</sup>Institute for Futures Studies, Box 591, SE-101 31 Stockholm, Sweden.

**Human communication in social networks is dominated by emergent statistical laws such as non-trivial correlations and temporal clustering. Recently, we found long-term correlations in the user's activity in social communities. Here, we extend this work to study the collective behavior of the whole community with the goal of understanding the origin of clustering and long-term persistence. At the individual level, we find that the correlations in activity are a byproduct of the clustering expressed in the power-law distribution of inter-event times of single users, i.e. short periods of many events are separated by long periods of no events. On the contrary, the activity of the whole community presents long-term correlations that are a true emergent property of the system, i.e. they are not related to the distribution of inter-event times. This result suggests the existence of collective behavior, possibly arising from nontrivial communication patterns through the embedding social network.**

Various constituents of social systems have been found to follow remarkable statistical regularities. Only the recent availability of relevant data made it possible to unravel such features. Tracking bank notes or cell phones it has been shown that humans follow simple and reproducible mobility patterns<sup>1,2</sup>. The communication via e-mails occurs in bursts, exhibiting a broad distribution of times between successive messages of individuals (inter-event times)<sup>3,4</sup>. Recently, we have found that the act of sending messages of individual users in two online communities present long-term correlations<sup>5</sup> characterized by power-law correlation functions obtained via standard Detrended Fluctuation Analysis.

In the present work we examine the relation between the two empirical findings of broad inter-event time distributions<sup>3,4</sup> and the long-term persistence identified in the communication activity<sup>5</sup>. Therefore, we investigate the communication activity of actants in a social online community with special consideration of the timing and study long-term correlations in the communication as well as clustering of successive messages. Here, the term *clustering* is used when the events tend to occur in burst, i.e. packages of many events are separated by long periods without events. In the case of power-law inter-event times this takes place on all scales. In other words, in the case of (temporal) clustering, the inter-event time distributions are more inhomogeneous than in the case of Poissonian statistics.

Long-term correlations have been found in the dynamics of many physical, technological, and natural systems. They are characterized by a divergent correlation time, i.e. a power-law decaying auto-correlation function (for a review see<sup>6</sup>). Such correlations lead to a pronounced mountain-valley-structure on all time scales – comprising indeterministic epochs of small and large values<sup>7</sup>. This type of persistence represents a surprising regularity since it is present in many different data such as DNA-sequences, human heartbeat, climatological temperature, etc.<sup>8–10</sup>. Long-term persistence in human related data has been reported for highway traffic<sup>11,12</sup>, Wikipedia access<sup>13</sup>, Ethernet traffic<sup>14</sup>, finance and economy<sup>15–17</sup>, written language<sup>18,19</sup>, as well as physiological records<sup>9,20,21</sup>. Human brain activity<sup>22–24</sup> and human motor activity<sup>25</sup> also comprise long-term correlations as well as city growth<sup>26–29</sup>, biological networks<sup>30</sup> and the spreading of disease<sup>31</sup>.

The distributions of inter-event times (times between successive messages) have been found to be rather broad, described by power-laws<sup>3</sup>. If many short intervals are separated by few long ones, the activity as messages per unit time comprises persistence, i.e. epochs of large and small activity. Since such distributions have been described with power-laws, we wish to investigate the relation between the long-term correlations in activity<sup>5</sup> and the broad



(power-law) distribution of inter-event times<sup>3</sup>. We will test two possible scenarios: (i) In the first scenario, the long-term correlations found in the communication activity<sup>5</sup> result from Levy type distributions, i.e. correlations are only due to the power-law inter-event time distribution (with exponents in the specific range)<sup>32</sup>. In the second scenario, (ii) the activity comprises ‘real’ correlations, i.e. the inter-event time distributions do not follow a power-law, but the communication activity is temporally not independent, namely long-term correlated.

We study the activity of sending messages based on detailed temporal data from a social online community and obtain the long-term correlation exponent  $H$  via DFA. The exponent  $H$  depends on the overall activity of the members; the more active the members the larger the fluctuation exponents. This exponents reaches a value  $H \approx 0.90$  for the most active users from an uncorrelated value  $H \approx 0.5$  for the less active ones. Then, we compare the value of  $H$  with the corresponding exponents of randomized data and a theoretical prediction relating correlations with clustering in the inter-event times. From the consistency of the comparison of this three measures, we conclude that the long-term correlations found in the activity of sending messages for single users is a direct consequence of the power-law distributed inter-event time of the individuals. Thus, the burstiness in the user activity explains the long-term correlations.

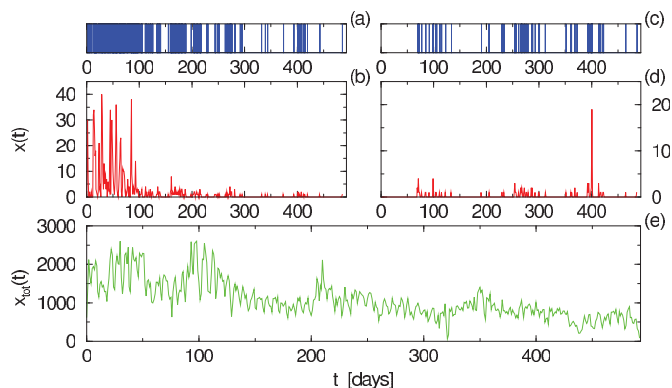
More interesting results are found when we consider the activity of the whole community as a sum of the activity of its members. Again we find non-trivial long-term correlations with exponents  $H$  in the same range as the individual users. However, the origin of this correlations is not related to the inter-event activity. This is probed by shuffling the activity data but preserving the distribution of inter-event times. In this case, this shuffling destroys the long-term correlations, implying that the correlations are not a byproduct of the broad distribution of inter-event times. We conclude that the whole system acts as a true long-term correlated system where correlations are not directly related to the Levy distributions of events.

We analyze the data of an online community (www.pussokram.com, POK<sup>33–35</sup>), covering the complete lifetime of the community over 492 days from February 2001 until June 2002. We record the activity among almost 30,000 members with more than 500,000 messages sent. This internet-site has been used for general social interactions and dating. The data consist of the time when the messages are sent and anonymous identification numbers of the senders and receivers. The data has been analyzed by us in<sup>5,36</sup>. In contrast to similar network data sets consisting only of snapshots, i.e. temporally aggregated social networks expressing who sent messages to whom, the advantage of this data set is that it provides the exact time when the messages were sent. For a discussion see<sup>37</sup>.

Before shutdown, the members could log in and meet virtually. In such communities, there are different ways of interacting. Usually, it is possible to choose favorites, i.e. certain members, that a person somehow feels committed to. Such platforms also offer the possibility to discuss in groups with other members about specific topics. We focus on messages sent among the members – they are similar to e-mails but have the advantage that they are sent within a closed community where there are no messages coming from or going outside. Figure 1 illustrates patterns of sending messages for typical single users [a–d] and for the whole community [e]. The data is publically available at [http://lev.cny.cuny.edu/~hmkse/soft\\_data.html](http://lev.cny.cuny.edu/~hmkse/soft_data.html). We would like to note that we do not consider here the QX dataset which we analyzed in<sup>5,36</sup>, since it covers only 2 months and the scaling of the distribution of inter-event times is not reliable and we could not measure the shape of this distribution consistently.

## Results

**Study of correlations in individual activity.** Applying DFA<sup>21,38,39</sup> we have found in<sup>5,36</sup> that the individual activity records,  $x(t)$ , i.e. messages per unit time (records of messages per day or per week),



**Figure 1 | Examples of activity of sending messages and overall activity in POK.** The vertical lines in (a) and (c) represent the instants when the messages have been sent by two arbitrary members. The panels (b) and (d) show the records of number of messages per day,  $x(t)$ , of the same two members. The record of the total number of messages sent by all members per day within POK is depicted in (e). (a) and (b): member326 ( $M = 1023$ ); (c) and (d): member9414 ( $M = 100$ ).

exhibit long-term correlations. The fluctuation function provided by DFA scales as

$$F(\Delta t) \sim (\Delta t)^H, \quad (1)$$

where the exponent  $H$  is also known as the Hurst exponent. In the case of long-term correlations – which are characterized by a power-law decaying auto-correlation function:

$$C(\Delta t) = \frac{1}{\sigma_x^2} \langle [x(t) - \langle x(t) \rangle][x(t + \Delta t) - \langle x(t) \rangle] \rangle \\ \sim (\Delta t)^{-\gamma},$$

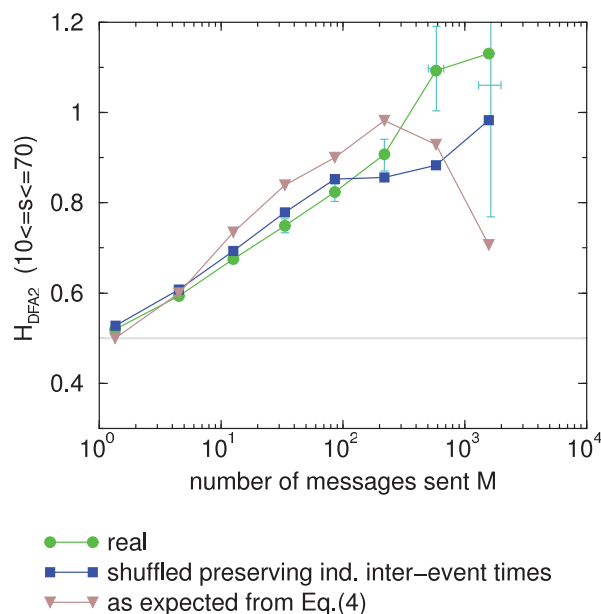
where  $\langle \cdot \rangle$  denotes the average,  $\sigma_x$  is standard deviation of  $x(t)$ , and  $\gamma$  is the correlation exponent ( $0 \leq \gamma \leq 1$ ) – one finds  $1/2 \leq H \leq 1$ , whereas larger exponents correspond to more pronounced long-term correlations. For uncorrelated or short-term correlated records ( $\gamma \geq 1$ , or in general  $\gamma \geq d$ ,  $d$  is the substrate dimension) the asymptotic fluctuation exponent is  $H = 1/2$ . In the range  $0 \leq \gamma \leq 1$  both exponents are related via

$$\gamma = 2 - 2H. \quad (2)$$

For an overview, we refer to<sup>6,39</sup>. DFA $n$  removes polynomial trends of the order  $n - 1$  from the original record  $x(t)$ , i.e. DFA2 copes with linear trends.

It is important to note that the DFA fluctuation function Eq. (1) is not applied to the activity  $x(t)$ , but to the integrated signal  $y(t) = \Sigma'x(t')$ . Thus,  $x(t)$  would be the analogous to the steps in a random walk and  $y(t)$  the displacement. DFA incorporates an additional detrending of the data. The integration leads to the appearance of long-term correlation when the interval between each step is power-law distributed. We will come back to this result when explaining the long-term correlations in terms of the burstiness.

We have measured the fluctuation exponents by applying least squares fits to  $\log F(\Delta t)$  vs.  $\log \Delta t$  on the scales  $10 < \Delta t < 70$  weeks conditional to the member’s activity level, e.g. their total number of messages,  $M^5$ . Figure 2 depicts the DFA results. We find that the less active members, sending very few messages in the period of data acquisition, exhibit uncorrelated behavior. The more messages the members send, the more correlated is their activity. The fluctuation exponent  $H$  increases with  $M$  and reaches values up to  $H = 0.91 \pm 0.04$  (value obtained for sending messages, we disregard the last points,  $M > 400$ , which have too large errors bars). The



**Figure 2 | Fluctuation exponents of the communication activity sending messages.** The exponents are plotted as a function of the activity level  $M$  (final number of messages) for the original data (green circles), shuffled data preserving the individual inter-event times (blue squares), as well as the exponents expected from Eq.(4) (brown triangles down) and from the distribution of inter-event times as characterized by power-law fits to the curves of Fig.4 providing the exponent  $\mu$ . The error bars were obtained from randomly separating the members into 10 groups and repeating the analysis for each of them.

uncorrelated behavior ( $H \approx 0.5$ ) for small activity can be understood since when  $M \approx 1-10$  there is not enough time in the data acquisition window to capture long-term correlations. Thus, the change from  $H = 0.5$  to  $H = 0.91$  might be most probably due to a crossover behavior due to finite acquisition time. In<sup>36</sup> we propose a model which reproduces the dependence of the fluctuation exponents on the activity level of the members. For receiving messages we find almost identical results<sup>36</sup>. We use weekly resolution in order to cope with possible weekly oscillations<sup>4,40-42</sup>.

Similar long-term correlations have been found in<sup>43,44</sup> in traded values of stocks and e-mail communication. The fluctuation exponent increases with the mean trading activity of the corresponding stock or with the average number of e-mails similarly as in our results.

**Study of clustering in individual activity.** The timing of human communication activity has been found to comprise bursts where many events occur in relatively short periods which are separated by long periods with few or no events at all. Such patterns can be characterized with the inter-event times, i.e. the times,  $dt$ , between successive messages. For e-mail communication it has been argued that their probability density follows a power-law,

$$P(dt) \sim (dt)^{-\mu}, \quad (3)$$

with exponent  $\mu \approx 1^{3,45,46}$ . As an origin for such heavy tails in human dynamics a queuing model has been suggested<sup>3</sup> according to which each individual performs tasks from a priority list. It has been confirmed that such a process can reproduce bursts of activity or clustering, see e.g.<sup>47,48</sup>. In contrast, analyzing the same e-mail data, a log-normal distribution has been found to be more appropriate to describe the inter-event time distribution<sup>49,50</sup>. We would like to remark that fitting fat tailed distributions is disputed<sup>51-54</sup>. There is neither a consensus on a typical functional form nor on a proper fitting technique. Recently, a cascading Poisson process based on

daily and weekly cycles has been proposed as origin of slower-than-exponential decays of  $P(dt)$ <sup>4,42</sup>. We studied the cascading Poisson process in<sup>36</sup>.

In<sup>55</sup>, memory in the sequences of  $dt$  has been studied for different data sets, characterizing the inter-event times in terms of a burstiness parameter, which is based on the distribution, and in terms of a memory coefficient, which is the auto-correlation function at lag 1. In addition, the authors locate the corresponding data sets in a phase diagram defined by these two quantities. Nevertheless, we would like to note that the quantification of long-term correlations in the  $dt$  can be hindered by noise<sup>56,57</sup>.

Next, we study the POK data, i.e. the inter-event times  $dt$  between successive messages of individual members, and relate their statistics to the long-term correlations. The finding of long-term correlations opens the question of the origin of such a persistence pattern in the social communication. From a statistical physics point of view, we consider two possible scenarios:

1. In the first scenario, the intervals between the messages follow a power-law<sup>3,58</sup>. Accordingly, the activity pattern comprises many short intervals and few long ones, implying persistent epochs of small and large activity. This fractal-like clustering in the activity can – depending on the exponent – lead to long-term correlations with  $H > 1/2$  (see the analogous problem of the origin of long-term correlations in DNA sequences as discussed in<sup>59</sup>). This scenario implies a direct link between the correlations in the activity and the distribution of inter-event times which can be obtained analytically<sup>60</sup>. We call this scenario “Levy correlations” since the actual activity may not be correlated per-se, but correlations arise as a byproduct of integrating a signal with a power-law distribution of inter-events in the DFA formalism.
2. In the second scenario, the intervals between the messages may or may not follow a power-law distribution, but the values of the inter-event times are not independent of each other and comprise ‘real’ long-term persistence. For example, the distribution of inter-event times could be stretched exponential (see recent work on the study of extreme events of climatological records exhibiting long-term correlations<sup>56,61</sup>) and then the only way to explain long-term correlations in the activity are correlations in the inter-event times. We call this scenario “true correlations” since the correlations are not related to the distribution of inter-events but they reflect ‘real’ correlations in the dynamics of the communication activity.

A possible way to discern between these two scenarios is to shuffle the temporal activity, keeping the inter-event distribution intact. While in the case of Levy type correlations shuffling the inter-event times should not influence the long-term correlation properties of  $x(t)$ , in the case of ‘real’ long-term correlations shuffling the inter-event times should destroy the (asymptotic) long-term correlations since the memory is due to the arrangement of the inter-event times. In what follows, we investigate the activity of individual members and the activity of the whole POK community.

**Study of inter-event distribution of individual members.** Figure 2 exhibits the fluctuation exponents for individual members when we shuffle the data but preserve the distribution of inter-event times. This is done according to the following steps: (i) Extract the set of inter-event times of each user. (ii) Shuffle the extracted data. (iii) Rebuild the record of events. Since the sum of the inter-event times does not add up to the entire period of data acquisition, the first event is chosen so that the remaining time is split into two, one part in the beginning and the other one at the end. (iv) Repeat the analysis.

The corresponding exponents also reach high values, almost as high as for the original data, and do not drop for very active members. This agreement is a first indication of Levy correlations in single user activity.



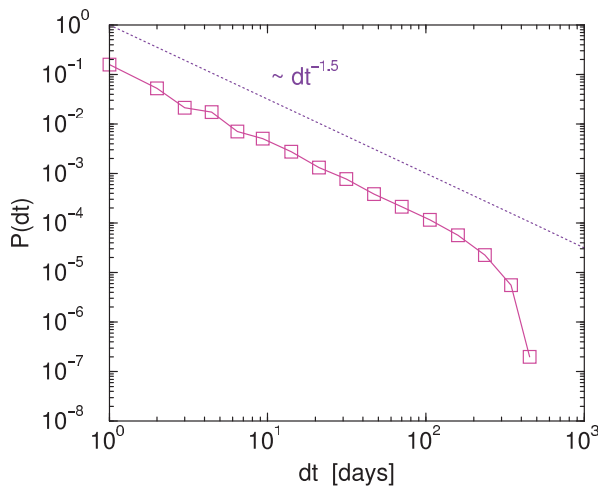
Further evidence is found by studying the distribution of inter-event times in the activity of each individual. Figure 3 shows the probability density,  $P(dt)$ , of times between messages of the same users sent in the online community. A power-law regime of approximately two decades can be seen with an exponent  $\mu \approx 1.5$ , which differs from the exponent reported for e-mail communication<sup>3,46</sup>, i.e.  $\mu \approx 1$ . A reason for these different findings might be that in the case of<sup>3</sup> only one user is considered and that  $\mu$  depends on the activity level of the users, as we show below. In addition, here we study all messages from a closed community. The exponent we find is closer to the one reported for reply times (waiting times), i.e. the time individuals spend between receiving and sending to the same communication partner. For reply times of e-mails and land mail  $\mu_w \approx 1.5$  has been reported<sup>3,62</sup>.

Since we found a dependence of the fluctuation exponent  $H$  on the activity level  $M$ , i.e. the total number of messages each member sends, we suspect that also  $\mu$  might depend on  $M$ . Thus, in Fig. 4 we plot for sending messages in POK (daily resolution) the  $P(dt)$  for groups of different activities, i.e. different total number of messages  $M$ . We find that for the most active members  $P(dt)$  decays rather steeply, while for the least active members  $P(dt)$  decays much slower. Due to the finite size of the data it is not quite clear which functional form the curves follow. If one assumes a power-law decay then the exponents are roughly in the range  $1 \leq \mu \leq 3$ .

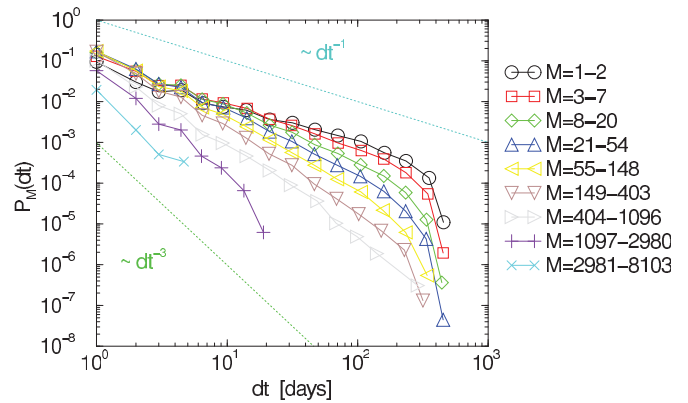
As discussed above, the power-law distribution of inter-event times, Eq. (3), can lead to long-term correlations in activity, without requiring temporal dependencies between the intervals themselves. It can be shown that the long-term persistence properties of this point process are characterized by the fluctuation exponent which theoretically depends on  $\mu$  according to<sup>23,32,60,63</sup>:

$$H_\mu = \begin{cases} \mu/2 & \text{for } 1 < \mu < 2 \\ 2 - \mu/2 & \text{for } 2 < \mu < 3, \\ 1/2 & \text{else} \end{cases} \quad (4)$$

see Fig. 5. Apart from detrending, DFA provides an integration of the original record. So if there are long periods of no activity due to power-law inter-event times, then, this is reflected in long-term persistence in the signal calculated by DFA. Thus, the existence of long-term correlations is due to the long periods distributed via Levy distributions as expressed by the direct relation between correlations and Levy inter-event activity, Eq. (4).



**Figure 3 | Probability density of inter-event times  $dt$  between successive messages sent by a single member of POK, in daily resolution.** The values are extracted by considering every single individual sending messages in the period of data acquisition and then joined from all members. The dotted line in the top corresponds to the exponent  $\mu = 1.5$ .

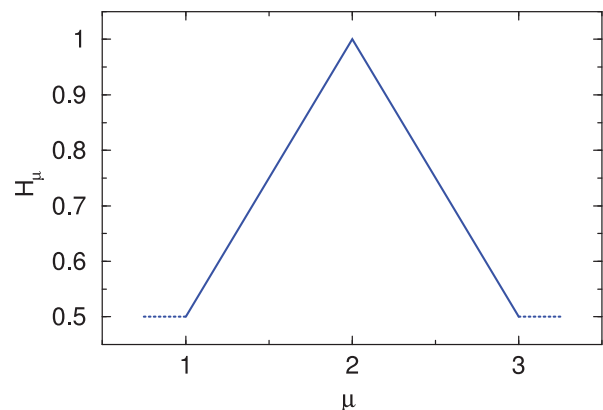


**Figure 4 | Probability density of inter-event times  $dt$  between successive messages sent by all individual members of POK in daily resolution.** The values are extracted for the individuals and unified among members according to their activity level  $M$ . The curve for the most active members is in the bottom, while the one for the least active is in the top. The dotted lines correspond to the exponents  $-1$  (top) and  $-3$  (bottom).

Applying least squares fits (in the straight range) to the  $P(dt)$  for sending in POK (Fig. 4) we obtain values for  $\mu$  as a function of the activity level  $M$  and determine the corresponding fluctuation exponents,  $H_\mu$ , as expected from Eq. (4). We would like to note that the curves in Fig. 4 are not always straight lines leading to large uncertainty regarding the estimated values of  $\mu$ .

Figure 2 depicts the fluctuation exponents  $H_\mu$  from Eq. (4) in comparison with the values obtained from DFA. We find  $H \approx H_\mu$  for a big part of the  $M$  range. The exponents  $H_\mu$  are also close to  $H$  of the shuffled records where the inter-event times are preserved. The fact that when we shuffle the signal, respecting the corresponding distribution of inter-event, gives rise to the same correlation function, indicates that the origin of the long-term correlation obtained in DFA are due to the Levy correlations. This is further corroborated by the agreement between  $H$  from DFA and the prediction  $H_\mu$ . From Fig. 2 we see that the three curves are in a reasonable agreement. This supports that the correlations in single user activity can be due to the power-law distribution of the inter-event times, which is in favor of Levy type correlations.

**Study of whole community activity.** Next, we investigate the activity of the community as a whole. While we have studied the activity of single users, it is of interest to investigate the activity of the whole community by considering the number of messages sent by all



**Figure 5 | Levy correlations and persistence.** Theoretical relation between the inter-event time distribution exponent  $\mu$  and the fluctuation exponent  $H_\mu$  according to Eq. (4).





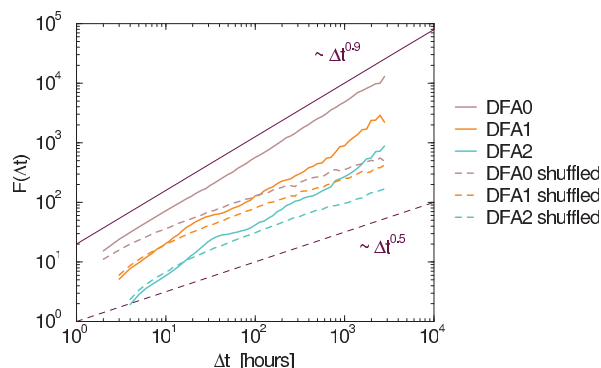
members in a specified period of time. Figure 1(e) shows such activity temporally aggregated to one day. The interest arises since we would like to test the existence of correlations emerging from collective behavior in the communication patterns at the level of the whole community.

For this study, we disregard who sends the messages to whom and only consider the instants when any message was sent. In order to have a sufficiently long record to apply DFA, we aggregate the data to messages per hour (instead of daily or weekly resolution). As can be seen in Fig. 1(e), the record contains oscillations<sup>4</sup>. Since such periodicities lead to erratic fluctuation functions<sup>39</sup>, we subtract the hourly averages over all days:  $x_{\text{tot}}(t) \rightarrow x_{\text{tot}}(t) - \langle x_{\text{tot}} \rangle_{t \bmod 24}$ .

The DFA fluctuation functions are shown in Fig. 6. The hump on scales around 20 hours in the results of DFA1 and DFA2 are residual oscillations, i.e. they were not completely removed. On larger scales this effect vanishes and we find a fluctuation exponent  $H_{\text{tot}} \approx 0.9$ . The straight line in the case of DFA0 is due to the fact that the maximum exponent is 1<sup>39</sup>. More importantly, when the record of the whole community is shuffled but preserving the inter-event distribution, the asymptotic scaling is  $F \sim (\Delta t)^{1/2}$ . That is, in contrast to the result for individual activity, when we shuffle the signal of the whole community, we obtain the uncorrelated exponent:  $H \approx 0.5$  (dashed lines in Fig. 6). The fact that the correlations vanish ( $H = 0.9 \rightarrow H = 0.5$ ) when the data is shuffled indicates that the long-term correlations found in the activity of the community as a whole are not due to Levy correlations. Instead, correlations in the whole community are “true correlations” appearing as a manifestation of collective behavior of the scale of the entire community.

Another surprise appears when we calculate the distribution of inter-event times for the whole community. Here we define inter-event the time between the sending of two consecutive messages of any member in the community. This contrasts with the same study done at the single user level (Fig. 4) when inter-event is defined as the time between two events of the same user. In a sense,  $P(dt)$  for the entire community captures the collective behavior emerging from the entire community as information travels through the network.

In Fig. 7 the resulting probability density is displayed. We find a plateau up to 50 seconds followed by a power-law decay according to Eq. (3) with  $\mu \approx 2.25$ . Thus, the distribution of inter-event activity of the community as a whole is also a Levy type like the single user activity, albeit with a larger exponent. Such a larger exponent reflects



**Figure 6** | Fluctuation function of the record of messages sent by any member of POK. The record is the same as in Fig. 1(e) but in hourly resolution. Prior to applying DFA, the record has been deseasoned according to  $x_{\text{tot}}(t) \rightarrow x_{\text{tot}}(t) - \langle x_{\text{tot}} \rangle_{t \bmod 24}$ . The different curves differ in the DFA-order (DFA0-DFA2, from top to bottom), which determines the capability of detrending. DFA2 eliminates linear trends in  $x_{\text{tot}}(t)$ <sup>39</sup>. The dotted line in the bottom corresponds to a power-law with exponent  $H = 0.5$  and serves as guide to the eye while the continuous line at the top represents  $H = 0.9$ .

the fact that  $P(dt)$  is narrower for the community than for the individuals, as expected.

When we convert the exponent  $\mu \approx 2.25$  to the  $H_{\mu}$  through the Levy distribution model, Eq. (4), we find  $H_{\mu} \approx 0.88$ . Thus surprisingly, Eq. (4) may also explain the persistence as in the individual activity. However, the main evidence of Fig. 6, that is, the fact that the correlations vanish when we shuffle the data, probe that, even if Eq. (4) provides a good estimation of  $H$ , the long-term correlations are due to ‘real’ correlations and are not an artifact of the integration of a Levy type activity with DFA.

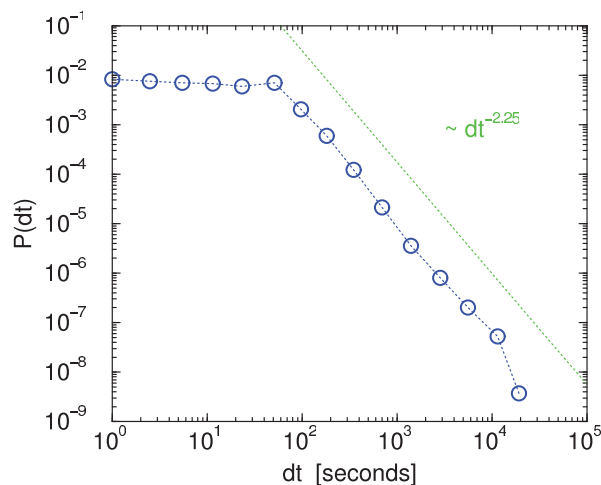
The long-term correlations found in the behavior of the entire community is more understandable than in the activity of single members, since the activity of the community is based on the communication patterns of the messages and information flowing through the whole system. The existence of  $H \approx 0.9$  at the whole level and the indications that the correlations are real ones is an interesting instance of the emergence of critical behavior in the collective dynamics of the system as a whole.

We conclude that while at the individual level we find Levy correlations, the activity of the whole community comprises ‘real’ correlations, which is due to the (possibly correlated) superposition of the individuals activity into a collective self-organized information flow in the system. Such a behavior is reminiscent of critical systems in phase transitions.

## Discussion

We have studied the timing of communication in a social online community and find long-term persistence in the activity of sending messages at the single user level and the whole community level. Furthermore, we have addressed the question of the origin of these long-term correlations and whether these are Levy type or ‘real’ correlations. While in the case of Levy type correlations the inter-event times need to be power-law distributed, ‘real’ long-term correlations are independent of the distributions, since they are due to interdependencies in the activities.

Our work, then, still leaves unanswered the question of the cause of the long-term persistence in the communication patterns at the whole community level. One possibility is that the temporal correlations are related to correlations in the network structure<sup>64,65</sup>. The persistence could also be due to social effects, i.e. the dynamics in the social network<sup>66</sup> induces persistent fluctuations, such as cascades. An example could be that a group of friends tries to make an appointment and therefore sends many subsequent messages in a relatively



**Figure 7** | Probability density of inter-event times  $dt$  between successive messages sent by any member of POK in seconds. The dotted straight line corresponds to a power-law with exponent  $\mu = -2.25$  and serves as guide to the eye.



short time<sup>67</sup>. After agreeing, the communication activity among the group drops. The activity patterns of individuals could be understood as a superposition of many such cascades. On the other hand, it could be purely due to a state of mind<sup>23</sup>, solipsistic, emerging from moods. More research is needed to thoroughly understand the interesting properties of human activity and its motives.

In conclusion, we have determined 3 exponents to characterize communication activity: (i)  $H$ , the fluctuation exponent of the original data, (ii)  $H_{\text{shuf}}$ , the fluctuation exponent when the data is shuffled preserving the inter-event times, (iii)  $H_{\mu}$ , the fluctuation exponent which is expected from power-law distributed inter-event times. We find that  $H \approx H_{\text{shuf}} \approx H_{\mu} \approx 0.9$  which supports the hypothesis of Levy correlations in the single user activity, while we find  $H \approx 0.9 \neq H_{\text{shuf}} \approx 0.5$  for the collective behavior of the whole community revealing non-trivial long-term correlations and self-organization at the level of the whole system.

We should mention a third scenario which we leave for future work. It is possible that the correlations comprise more complex features. It has been shown that nonlinear correlations in multifractal data sets lead to power-law distributed inter-event times (of peaks over threshold)<sup>68</sup>. In fact, the authors of<sup>68</sup> find in their Fig. 1(c) a similar dependence of  $\mu$  on the total number of events as we do for  $H_{\mu}$  in our Fig. 2. Additional analysis is needed to fully characterize the multifractal properties<sup>69–71</sup> of communication activity via e-mails or messages in online communities.

- Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
- Gonzalez, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
- Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
- Malmgren, R. D., Stouffer, D. B., Motter, A. E. & Amaral, L. A. N. A Poissonian explanation for heavy tails in e-mail communication. *Proc. Nat. Acad. Sci. U.S.A.* **105**, 18153–18158 (2008).
- Rybski, D., Buldyrev, S. V., Havlin, S., Liljeros, F. & Makse, H. A. Scaling laws of human interaction activity. *Proc. Nat. Acad. Sci. U.S.A.* **106**, 12640–12645 (2009).
- Kantelhardt, J. W. *Encyclopedia of Complexity and System Science*, chapter entry 00620: Fractal and Multifractal Time Series. Springer, 2009.
- Makse, H. A., Havlin, S., Schwartz, M. & Stanley, H. E. Method for generating long-range correlations for large systems. *Phys. Rev. E* **53**, 5445–5449 (1996).
- Peng, C.-K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Sciortino, F., Simons, M. & Stanley, H. E. Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170 (1992).
- Peng, C.-K., Mietus, J., Hausdorff, J. M., Havlin, S., Stanley, H. E. & Goldberger, A. L. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Phys. Rev. Lett.* **70**, 1343–1346 (1993).
- Koscielny-Bunde, E., Bunde, A., Havlin, S., Roman, H. E., Goldreich, Y. & Schellnhuber, H.-J. Indication of a universal persistence law governing atmospheric variability. *Phys. Rev. Lett.* **81**, 729–732 (1998).
- Tadaki, S., Kikuchi, M., Nakayama, A., Nishinari, K., Shibata, A., Sugiyama, Y. & Yukawa, S. Power-law fluctuation in expressway traffic flow: Detrended fluctuation analysis. *J. Phys. Soc. Jpn.* **75**, 034002 (2006).
- Xiao-Yan, Z., Zong-Hua, L. & Ming, T. Detrended fluctuation analysis of traffic data. *Chin. Phys. Lett.* **24**, 2142–2145 (2007).
- Kämpf, M., Tismer, S., Kantelhardt, J. W. & Muchnik, L. Burst event and return interval statistics in wikipedia access and edit data. *submitted* (2011).
- Leland, W. E., Taqqu, M. S., Willinger, W. & Wilson, D. V. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. Networking* **2**, 1–15 (1994).
- Liu, Y., Gopikrishnan, P., Cizeau, P., Meyer, M., Peng, C.-K. & Stanley, H. E. Statistical properties of the volatility of price fluctuations. *Phys. Rev. E* **60**, 1390–1400 (1999).
- Mantegna, R. N. & Stanley, H. E. *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).
- Lux, F. & Ausloos, M. *The Science of Disasters*, chapter 13. Market Fluctuations I: Scaling, Multiscaling, and Their Possible Origins, pages 373–409 (Springer-Verlag, Berlin, 2002).
- Schenkel, A., Zhang, J. & Zhang, Y.-C. Long range correlations in human writings. *Fractals* **1**, 47–57 (1993).
- Kosmidis, K., Kalampokis, A. & Argyrakis, P. Language time series analysis. *Physica A* **370**, 808–816 (2006).
- Ivanov P., Ch., Bunde, A., Amaral, L. A. N., Havlin, S., Fritsch-Yelle, J., Baevsky, R. M., Stanley, H. E. & Goldberger, A. L. Sleep-wake differences in scaling

behavior of the human heartbeat: Analysis of terrestrial and long-term space flight data. *EPL* **48**, 594–600 (1999).

- Bunde, A., Havlin, S., Kantelhardt, J. W., Penzel, T., Peter, J.-H. & Voigt, K. Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Phys. Rev. Lett.* **85**, 3736–3739 (2000).
- Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M. & Ilmoniemi, R. J. Long-range temporal correlations and scaling behavior in human brain oscillations. *J. Neurosci.* **21**, 1370–1377 (2001).
- Allegrini, P., Menicucci, D., Bedini, R., Fronzoni, L., Gemignani, A., Grigolini, P., West, B. J. & Paradisi, P. Spontaneous brain activity as a source of ideal  $1/f$  noise. *Phys. Rev. E* **80**, 061914 (2009).
- Gallos, L. K., Makse, H. A. & Sigman, M. A small-world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proc. Nat. Acad. Sci. USA* **109**, 2825–2830 (2012).
- Ivanov P., Ch., Hu, K., Hilton, M. F., Shea, S. A. & Stanley, H. E. Endogenous circadian rhythm in human motor activity uncoupled from circadian influences on cardiac dynamics. *Proc. Nat. Acad. Sci. U.S.A.* **104**, 20702–20707 (2007).
- Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
- Makse, H. A., Andrade, J. S., Batty, M., Havlin, S. & Stanley, H. E. Modeling urban growth patterns with correlated percolation. *Phys. Rev. E* **58**, 7054–7062 (1998).
- Rozenfeld, H. D., Rybski, D., Andrade, J. S. Jr., Batty, M., Stanley, H. E. & Makse, H. A. Laws of population growth. *Proc. Nat. Acad. Sci. USA* **105**, 18702–18707 (2008).
- Rozenfeld, H. D., Rybski, D., Gabaix, H. A. & Makse, X. The area and population of cities: New insights from a different perspective on cities. *American Economic Review* **101**, 2205–2225 (2011).
- Galvao, G., Miranda, J. G. V., Andrade, R. F. S., Andrade, J. S. Jr., Gallos, L. K. & Makse, H. A. Modularity map of the network of humn cell differentiation. *Proc. Nat. Acad. Sci. USA* **107**, 5750–5755 (2010).
- Gallos, L. K., Barttfeld, P., Havlin, S., Sigman, M. & Makse, H. A. Collective behavior in the spatial spreading of obesity. *Sci. Rep.* **2**, 454 (2012).
- Shlesinger, M. F., West, B. J. & Klafter, J. Lévy dynamics of enhanced diffusion: Application to turbulence. *Phys. Rev. Lett.* **58**, 1100–1103 (1987).
- Holme, P. Network dynamics of ongoing social relationships. *EPL* **64**, 427–433 (2003).
- Holme, P., Liljeros, F., Edling, C. R. & Kim, B. J. Network bipartivity. *Phys. Rev. E* **68**, 056107 (2003).
- Holme, P., Edling, C. R. & Liljeros, F. Structure and time evolution of an internet dating community. *Soc. Networks* **26**, 155–174 (2004).
- Rybski, D., Buldyrev, S. V., Havlin, S., Liljeros, F. & Makse, H. A. Communication activity in social networks: growth and correlations. *Eur. Phys. J. B* **84**, 147–159 (2011).
- Gallos, L. K., Rybski, D., Liljeros, F., Havlin, S. & Makse, H. A. How people interact in evolving online affiliation networks. *Phys. Rev. X2*, *in press* (2012).
- Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E. & Goldberger, A. L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* **49**, 1685–1689 (1994).
- Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S. & Bunde, A. Detecting long-range correlations with detrended fluctuation analysis. *Physica A* **295**, 441–454 (2001).
- Golder, S., Wilkinson, D. M. & Huberman, B. A. Rhythms of social interaction: messaging within a massive online network. *online-arXiv* (arXiv:cs/0611137v1 [cs.CY], 2006).
- Leskovec, J. & Horvitz, E. Planetary-scale views on an instant-messaging network. *online-arXiv* (arXiv:0803.0939v1 [physics.soc-ph], 2008).
- Malmgren, R. D., Stouffer, D. B., Campanharo, A. S. L. O. & Amaral, L. A. N. On universality in human correspondence activity. *Science* **325**, 1696–1700 (2009).
- Eisler, Z. & Kertész, J. Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. *Phys. Rev. E* **73**, 046109 (2006).
- Eisler, Z., Bartos, I. & Kertész, J. Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142 (2008).
- Johansen, A. Probing human response times. *Physica A* **338**, 286–291 (2004).
- Johansen, A. Comment on A.-L. Barabasi, Nature 435 207–211 (2005). *online-arXiv* (arXiv:physics/0602029v1 [physics.soc-ph], 2006).
- Vázquez, A. Exact results for the Barabasi model of human dynamics. *Phys. Rev. Lett.* **95**, 248701 (2005).
- Vázquez, A., Oliveira, J. G., Dezső, Z., Goh, K. I., Kondor, I. & Barabási, A.-L. Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E* **73**, 036127 (2006).
- Stouffer, D. B., Malmgren, R. D. & Amaral, L. A. N. Comment on “The origin of bursts and heavy tails in human dynamics” by Barabasi, Nature 435, 207 (2005). *online-arXiv* (arXiv:physics/0510216v1 [physics.data-an], 2005).
- Barabási, A.-L., Goh, K.-I. & Vazquez, A. Reply to comment on “the origin of bursts and heavy tails in human dynamics”. *online-arXiv* (arXiv:physics/0511186v1 [physics.data-an], 2005).
- Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
- Gabaix, X. & Ibragimov, R. Log(rank-1/2): a simple way to improve the ols estimation of tail exponents. Discussion Paper 2106 (26), Harvard Institute of Economic Research, Cambridge, Massachusetts, February (2006).
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).



54. Malevergne, Y., Pisarenko, V. & Sornette, D. Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal. *online-arXiv* (arXiv:0909.1281v1, 2009).
55. Goh, K.-I. & Barabási A.-L. Burstiness and memory in complex systems. *EPL* **81**, 48002 (2008).
56. Eichner, J. F., Kantelhardt, J. W., Bunde, A. & Havlin, S. Statistics of return intervals in long-term correlated records. *Phys. Rev. E* **75**, 011128 (2007).
57. Lennartz, S. & Bunde, A. Eliminating finite-size effects and detecting the amount of white noise in short records with long-term memory. *Phys. Rev. E* **79**, 066101 (2009).
58. Gerstein, G. L. & Mandelbrot B. Random walk models for spike activity of single neuron. *Biophys. J.* **4**, 41–68 (1964).
59. Buldyrev, S. V., Goldberger, A. L., Havlin, S., Peng, C.-K., Simons, M. & Stanley, H. E. Generalized Lévy-walk model for DNA nucleotide sequences. *Phys. Rev. E* **47**, 4514–4523 (1993).
60. Buldyrev, S. V. *Encyclopedia of Complexity and System Science* (volume Fractals and multifractals, chapter Fractals in Biology. Springer, 2010).
61. Bunde, A., Eichner, J. F., Kantelhardt, J. W. & Havlin, S. Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys. Rev. Lett.* **94**, 048701 (2005).
62. Oliveira, J. G. & Barabási, A.-L. Darwin and Einstein correspondence patterns. *Nature* **437**, 1251–1251 (2005).
63. Thurner, S., Lowen, S. B., Feurstein, M. C., Heneghan, C., Feichtinger, H. G. & Teich, M. C. Analysis, synthesis, and estimation of fractal-rate stochastic point processes. *Fractals* **5**, 565–595 (1997).
64. Kentsis, A. Mechanisms and models of human dynamics. *Nature* **441**, E5–E5 (2006).
65. Rybski, D., Rozenfeld, H. D. & Kropp, J. P. Quantifying long-range correlations in complex networks beyond nearest neighbors. *EPL* **90**, 28002 (2010).
66. Stehle, J., Barrat, A. & Bianconi, G. Dynamical and bursty interactions in social networks. *Phys. Rev. E* **81**, 035101 (2010).
67. Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).
68. Bogachev, M. I., Eichner, J. F. & Bunde, A. Effect of nonlinear correlations on the statistics of return intervals in multifractal data sets. *Phys. Rev. Lett.* **99**, 240601 (2007).
69. Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A. & Stanley, H. E. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A* **316**, 87–114 (2002).
70. Kantelhardt, J. W., Rybski, D., Zschiegner, S. A., Braun, P., Koscielny-Bunde, E., Livina, V., Havlin, S. & Bunde, A. Multifractality of river runoff and precipitation: comparison of fluctuation analysis and wavelet methods. *Physica A* **330**, 240–245 (2003).
71. Kantelhardt, J. W., Koscielny-Bunde, E., Rybski, D., Braun, P., Bunde, A. & Havlin, S. Long-term persistence and multifractality of precipitation and river runoff records. *J. Geophys. Res.-Atmos.* **111**, D01106 (2006).

## Acknowledgments

We thank C. Briscoe, J.F. Eichner, L.K. Gallos, and H.D. Rozenfeld for useful discussions. This work was supported by National Science Foundation Grants NSF-SES-0624116 and NSF-EF-0827508 and ARL F.L. acknowledges financial support from The Swedish Bank Tercentenary Foundation. S.H. thanks the European EPIWORK project, the Israel Science Foundation, ONR and DTRA for financial support.

## Author contributions

All authors contributed equally to the work presented in this paper including ideas, manuscript preparation and analysis.

## Additional information

**Competing financial interests:** The authors declare no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

**How to cite this article:** Rybski, D., Buldyrev, S.V., Havlin, S., Liljeros, F. & Makse, H.A. Communication activity in a social network: relation between long-term correlations and inter-event clustering. *Sci. Rep.* **2**, 560; DOI:10.1038/srep00560 (2012).