Karsai, Márton; Kaski, Kimmo; Kertész, János

**Correlated Dynamics in Egocentric Communication Networks**

*Published in:*
PloS one

*DOI:*
10.1371/journal.pone.0040612

Published: 01/01/2012

*Document Version*
Publisher's PDF, also known as Version of record

*Please cite the original version:*
Correlated Dynamics in Egocentric Communication Networks

Márton Karsai1*, Kimmo Kaski1, János Kertész1,2

1 BECS, Aalto University School of Science, Espoo, Finland, 2 Institute of Physics, Budapest University of Technology and Economics, Budapest, Hungary

Abstract

We investigate the communication sequences of millions of people through two different channels and analyse the fine-grained temporal structure of correlated event trains induced by single individuals. By focusing on correlations between the heterogeneous dynamics and the topology of egocentric networks we find that the bursty trains usually evolve for pairs of individuals rather than for the ego and his/her several neighbours, thus burstiness is a property of the links rather than of the nodes. We compare the directional balance of calls and short messages within bursty trains to the average on the actual link and show that for the trains of voice calls the imbalance is significantly enhanced, while for short messages the balance within the trains increases. These effects can be partly traced back to the technological constraints (for short messages) and partly to the human behavioural features (voice calls). We define a model that is able to reproduce the empirical results and may help us to understand better the mechanisms driving technology mediated human communication dynamics.

Introduction

Egocentric networks consist of a central node (the ego) and its immediate neighbours. They are broadly used in psychology and sociology as they are crucial in understanding the interactions between an individual and his/her proximate social circle [1,2]. Dunbar discovered that the number of social ties of an individual is anthropologically determined to be around 150 [3–5] and that the ties can be classified into self-containing intimacy circles [6] based on the strength of one’s relationships. These differences in the importance of the ties of an ego are reflected by the dynamics of social relationships. An individual interacts heavily only with very few people like close kins or closest friends and communicates less with most of his/her acquaintances depending on their social closeness and common tasks [7,8]. This diversity is naturally present also in the communication dynamics thus data collecting communication sequences of individuals can even be used to measure the intensity of relationships [10,11]. Fig. 1.a illustrates the communication pattern in an egocentric network, where the overall activity of the ego (green row) and activities on separated edges with three friends (orange rows) are presented. After building up the complete social network from egocentric subgraphs it turns out that strong links with heavy communication are usually found inside densely connected groups or communities as those people who are connected through an important relationship share many common neighbours. At the same time weak links more likely connect communities [9–11], which makes them important for processes evolving in social networks [12].

Individual human behaviour shows heterogeneity not only in topology but also in dynamics. The actions of a person are not evenly distributed but rather clustered in time (e.g. see Fig. 1.a). It commonly happens that several events are executed in bursts within a short time frame and such high activity periods are separated by long inactive ones. This kind of temporal inhomogeneity can be observed in various human activity sequences ranging from communication to library loans or printing logs [13–16] and they are typically characterized by the broad distribution of inter-event times $P(t)\sim t^{-\alpha}$. In the past few years different mechanisms were proposed to describe the origin of these heterogeneities [17–21], and several influences have been analyzed, including the role of circadian patterns [22].

Even though the inter-event time distribution assigns temporal inhomogeneities, it does not characterize possible correlations in the sequences. It was observed that such correlations between bursty events, indeed, exist as indicated by trains of bursty events in Fig. 1.b–d, where the incoming and outgoing calling activity of an individual are shown on each of his/her links for some selected time periods. By appropriately defining such correlated trains it turned out that the distribution of the number $E$ of events they contain shows a $P(E)\sim E^{-\beta}$ scaling [23]. This behaviour has been found in various human communication sequences and even in earthquake and neuron firing statistics, indicating same kind of universality hidden behind these correlated heterogeneous dynamics. These phenomena were interpreted as a result of memory effects and modelled [23] by simple multi-scale reinforcement processes similar to that describing face-to-face interactions [24,25].

In spite of some related works [21] the fine-grained structure of correlated trains in communication sequences are still mainly unexplored. In this paper by using detailed electronic records of large number of individuals we aim to study this problem and addressing the following questions: How are the event trains correlated with the skeleton of the backgrounding social network?
Are they results of collective behaviour or are they induced by individuals? Are there any correlations between the directions of the interactions? This paper is organized as follows. After introduction we define the basic quantities and show that the correlated bursty communication pattern can be attributed to links rather than to nodes. In the subsection below we compare the balance of in- and outgoing calls and SMS-s on a link to that of the trains on the same link. We then present a model study to gain deeper insight into the governing rules of the observed behaviour. After drawing the conclusions we close the paper by the detailed description of the investigated datasets and used methods.

Results

Dynamics in Egocentric Networks

In [23] it has been demonstrated that the correlations in a bursty time series showing a power-law inter-event time distribution $P(t_{ev})$ [14–16] can be detected by looking for bursty event trains and investigating the distribution $P(E)$ of the number $E$ of events they include. A train of bursty events consists of consecutive events all with inter-event times $t_{ev} \leq \Delta t$ inside the train, separated from the rest by $t_{ev} > \Delta t$. For any independent $P(t_{ev})$ (including power law ones) $P(E)$ will decay exponentially while for correlated signals a power-law behaviour of $P(E)$

$$P(E) \sim E^{-\beta}$$

appears and indicate strong correlation of bursty events. In the present case for voice calls (SMS) both $P(t_{ev})$ and $P(E)$ are distributed like a power-law with exponents $\gamma \geq 0.7$ ($\gamma \approx 1.0$) and $\beta = 4.2$ ($\beta = 3.9$) (see original solid lines in Fig. 2). This behaviour is remarkably different from the $P(E)$ in sequences where inter-event times of the whole data set are randomly shuffled. This procedure leaves the inter-event time distribution and the number of communication events between pairs of users unaltered but destroys all temporal correlations. Correspondingly, the $P(E)$ distribution decays exponentially (random solid lines on Fig. 2.a and b).

If we assume burstiness to be a mere consequence of individual human behaviour, it is natural to concentrate on outgoing voice calls and SMS’s. However, beyond studying only the timings of events, we can look at how they are distributed in the egocentric topology. Intuitively one may assume that correlated outgoing calls or SMS’s of an individual serve the information processing of a group [27,28], thus these events are directed towards several neighbours leading to the evolution of larger temporal motifs. The existence of such behaviour was demonstrated in [29] and also in Fig. 1.c where beyond the ego two other friends are involved in some bursty sequences. Supposing this mechanism be relevant then the burstiness would be the property of a single node or a group of individuals.

Surprisingly, the generic picture seems to be different. If we assume that the correlated events in a train are directed toward several neighbors, trains of events on single edges between two persons should have an entirely different, less correlated statistics of bursts. However, this is not the case as the $P(E)$ distributions which were detected on single edges are scaling similarly and can be characterized by the same exponent values as in the case, when calls on any egocentric edges were taken into account. This suggests that trains of events usually evolve on single links. Such behaviour was confirmed both for calls and SMS’s (Fig. 2.a and b), where the $P(E)$ distributions of trains evolving between pair of individuals (original solid lines) are only slightly different from the $P(E)$ distributions of trains which can involve several people (original dashed lines). This picture is also supported by the statistics of temporal motifs [29], where motifs involving two individuals are by far the most common ones.

The same conclusion can be drawn by counting the number of neighbours $BN$, whom an individual called in a bursty train of $E$ events. If a user communicates with only one neighbour during the
Figure 2. Characteristic distributions of call and SMS sequences. (a) Distributions of $E$ length of bursty periods of outgoing events of nodes towards all the neighbours (solid lines) and to the direction of single neighbours (dashed lines). The measurement was repeated with time windows $\Delta t = 100, 300, 600$ and 3600 seconds using the original event sequence (original) and after the inter-event times were shuffled (random). Inset: Inter-
period then the ratio $BN/E = 1/E$ or, e.g., if each calls are
directed toward different neighbours than $BN/E = 1$. The
distributions of the $BN/E$ ratios for each $E$ together with the
average values are presented in Fig. 3.a and b for calls and SMS,
respectively. For shorter trains the distributions show some
dispersion, however the mean values confirm that usually only
one or two persons are called in a bursty train. To estimate what
fraction of trains of the same size involves one, two or more
people, we calculate the cumulative fraction of trains involving $BN$
number of neighbours. As it is depicted in Fig. 3.c and d, for any $E$
the majority of the trains of events involve only one neighbour (red
areas) while the fractions with multiple call receivers are
considerably smaller. In addition Fig. 3.e and f indicate that even
though trains consist of events executed with several friends, most
of the calls are performed with only one of them. This can be
concluded as the average of the maximum number of calls
performed with the same neighbour in a train with size $E$ goes with
the train size as $(\text{max}(CN))(E) \sim E$.

In order to see the significance of the observed behaviour we
repeated the measurements on null model sequences where we
shuffled the receivers between calls of a single individual and repeated
the method for everyone in the dataset. This way we kept the
original degree of the users and the temporality of events remained
also the same. Consequently in the null model sequences we could
detect the same bursty trains as originally, however, since the
receivers were shuffled the trains did not evolve between the ego
and a single neighbour but were possibly distributed towards the
other neighbours of the actual user. This way we shuffled out the
effect of correlations causing the evolution of bursty trains on
single links. Results for call and SMS null model sequences are
presented in the Supplementary Informations (SI), where we show
that in the random case the fraction of single link bursty trains turn
to be around $\sim 20\%$ (which is $\sim 80\%$ in the original sequences)
while the fractions of trains evolving between the ego and larger
$BN$ number of friends are increasing accordingly. Hence single
link bursty train evolution is a relevant feature of human
communication behaviour.

Mutual Bursty Behaviour

In the previous section we pointed out similarities in the dynamics
of communication between calls and SMS. In both cases the
temporal distributions of events are strongly heterogeneous, the
sequences consist of correlated actions which are clustered into long
bursty trains, and which trains are usually evolving on single links
rather then connecting a larger group of people. Now we are at the
position to ask further questions about the dynamics of human
interactions on a dyadic level like: What is the direction of events in
the trains? Is the communication in bursty periods balanced or
dominated by one partner? Are there differences between voice calls
and SMS from this point of view? To answer these questions we take
into account the communication events performed on links
initiated by any of the two connected users.

For the entire recording period of the dataset one can calculate the
overall communication balance for each edge $e$ as:

$$b_e = \max(N_A,n_B)/N_e$$  (2)

where $N_A (N_B)$ assigns the total number of calls from $A$ to $B$ ($B$ to
$A$). Hence $b_e$ can vary between $1/2$ (completely balanced) and $1$
(completely imbalanced, dominated by one of the participants). Strong
difference occurs between voice call and SMS dynamics if we look at the balance of edges and their correlation with the
length of trains which evolve on them. To do so, we calculate the
weighted average of $b_e$ over all trains of length $E$:

$$b_{\text{all}}(E) = \langle b_e \rangle_{\text{trains}} = \frac{\sum_e n_e(E)b_e}{\sum_e n_e(E)}$$  (3)

where $n_e(E)$ means the number of trains of length $E$ on edge $e$. In
Fig. 4.a for SMS (brown squares) $b_{\text{all}}$ is converging to $1/2$
indicating that the trains evolve on strongly balanced links and this
effect is enhancing for longer trains. The balanced communication
in SMS sequences has been observed by Wu. et al. [21] and can be
explained by the technologically determined way of communica-
tion exchange, which requires mutuality for SMS. However, this
constraint does not apply for the mobile calls (orange points in
Fig. 4.a) where during a call information can flow in both directions.
Here $b_{\text{all}}$ reflects strongly unbalanced communication and it increases towards $1$ for trains with larger $E$. It assigns
an opposite trend for calls compared to SMS communication as
longer trains evolve on more unbalanced edges.

The next question we should address is whether the balance or
imbalance within the trains simply reflect the bias due to $b_e$ of the
actual edge or additional effects play role in there. To study this we
define the balance within a train in a similar way as for an edge:

$$p_e(E_m) = \max(N_A,n_B)/N_A + N_B.$$  (4)

Here $p_e(E_m)$ is the balance of the $m$-th train of length $E_m$ on
edge $e$ connecting $A$ and $B$, $n_g (n_B)$ denotes the number of
events initiated by $A$ ($B$) towards $B$ ($A$) in that train; $E_m = n_A + n_B$.
Averaging over trains of the same size $p(E) = \langle p(A,E_m) \rangle_{E\text{ trains}}$
gives an estimate for the average communication balance in trains of
size $E$ (note that in this case different $p_e(E_m)$ values can evolve
even for trains on the same edge $e$). This has to be compared to the
case, when a train is composed from events selected in an
independent manner from a set with balance $b_e$. The latter can be
Calculated as

$$p_e^{\text{indp}}(E_m = E) = \frac{1}{E} \sum_{i=0}^{E} \left( \frac{E}{2} - i \right) \left( \frac{E}{2} + i \right) b_e(1-b_e)^{E-i} \left( E \right),$$  (5)

where the first factor after the summation weighting the binomial
distribution taking into account that the imbalance can evolve in
either direction, i.e., parallel or antiparallel to the imbalance of the
edge. As $p_e^{\text{indp}}(E_m = E)$ depends only on $b_e$ and $E$, the average
can be taken similarly to $p(E)$ as $p_e^{\text{indp}}(E) = \langle p_e^{\text{indp}}(E_m) \rangle_{E\text{ trains}}$ to get
the estimate for the independent case.

Fig. 4.b shows $p(E)$ and $p^{\text{indp}}(E)$ for both voice calls and SMS
messages. The first apparent feature is that for large values of $E$
the $b_{\text{all}}(E)$ of Fig. 4.a and $p^{\text{indp}}(E)$ become very similar, correspond-
ing to the fact that for long trains $p^{\text{indp}}(E)$ approaches the bias of
the actual edge. However, the interesting effect is the difference
between $p(E)$ and the corresponding. It shows that trains of calls
(red points) are much more unbalanced than one would expect
from independent processes (yellow circles) considering the bias of
the edge. At the same time for SMS the contrary is true as trains (blue squares) are much more balanced than one would derive from random processes (green squares) applying the $b_e$ values of the corresponding edges. This demonstrates real correlation between events of the same train and suggests different correlated mechanisms behind call and SMS dynamics. One may suspect

Figure 3. Measures of single link bursty trains. (a, b) The distribution and average values of the ratio $BN/E$ ($BN$ being the number of neighbours, whom an individual called in a bursty train) for each $E$ train size. Trains were detected with $\Delta t = 600$ ($\Delta t = 300$) for calls (SMS). The pointed and solid lines denote the limiting case 1 and $1/E$, while dashed line belongs to $2/E$. (c, d) Cumulative fraction of the number of bursty neighbours ($BN$) for trains with size $E$ for calls and SMS. (e, f) Average of the maximum number of events directed to the same user within a bursty period in case of calls and SMS. doi:10.1371/journal.pone.0040612.g003
that the strengthening of the imbalance for voice calls may be a consequence of call trials picked up by the answering machine. To exclude this we repeated the calculations after removing events with duration shorter than $d$ as depicted in Fig. 4.c. The calculated $\text{pdf}_{\text{diff}}(E)$ in Fig. 4.d shows that the saturation value is slightly decreasing as we remove longer calls, however, the differences remain significant and saturating to $0:1$ with $d > 4$ seconds. This duration time also serves as a time scale after which the answering machine calls do not play any significant role. For further details about the effect of short calls see SI.

Model

In this section we introduce a model, which is able to reproduce the empirical observations, namely the enhanced imbalance of the bursty trains for voice calls and the opposite behaviour for the SMS message compared to independent processes. Our aim is to identify the different mechanisms controlling the dynamics of communication through these channels, integrate them into a single model which we can test against the empirical results.

The emergence of correlated bursty trains in communication sequences were interpreted earlier by memory processes [23]. The evolution of trains with size scaling as Eq.1 were explained as the result of reinforcement dynamics where the probability $q(n)$ to perform one more event in a train after $n$ events have been executed, depends on $n$ as:

$$ q(n) = \left( \frac{n}{n+1} \right)^\beta $$

where $\beta = 1$.

Assuming this dynamics and by fitting $\beta$ to empirical data, we are able to generate model trains with realistic size scaling. However, the question remains how we can introduce the mechanisms responsible for the enhancement of bursty trains.

Let us first concentrate on voice calls. One correlation we observed (see Fig. 4.a lower panel) is that longer trains tend to be more unbalanced, meaning that they are more dominated by one of the callers. We also disclosed the possibility (in Fig. 4.b and SI) that this behaviour is an artifact due to answering-machine calls where the caller repetitively recalls the friend to pass an important information. Keeping in mind that mobile calls enable bidirectional information change, we assume that the observed unbalanced communication in call trains reflects the difference in motivation between the communicating partners. If there is a task to solve, which is more important for one party, it gives motivation for him/her to repeated calls until the issue gets settled.
This mechanisms can be incorporated into the reinforcement process of bursty trains in the following way (see Fig. 3.a). We simulate bursty trains, which evolve on a link between a pair of individuals $A$ and $B$. To initiate a train with a probability equal to $b_0$ we randomly select $A$ or $B$ who then perform one event towards the other agent and at the same time we set the actual train size to $n = 1$. The decision about the next event is carried out in two steps. First we decide with the probability in Eq.6 whether to perform one more event in the train or initiate a new train otherwise. If the train should be continued a new decision is made about the direction of the call. The probabilities that it is initiated by $A$ or $B$ are given as:

$$q_a(n|\sigma_1) = \frac{n}{n+1} \quad \text{or} \quad q_b(n|\sigma_1) = 1 - \frac{n}{n+1} \quad (7)$$

where $\sigma \in \{A,B\}$. Here $q_a(n|\sigma_1)$ denotes the probability that the $n$th event of the actual train is performed by the same user who initiated the train at $n=1$, while $q_b(n|\sigma_1)$ gives the probability that the other player makes the call. Consequently, the longer a train evolves, the larger is the probability that the agent, who initiated the actual train will make a call to the other agent. Eq.6 and Eq.7 capture the coupled reasons for the evolution of long unbalanced trains. They reflect the reinforced motivation of an individual induced by the effort what he/she already invested in the actual series of calls to successfully solve a task with the other partner.

In case of SMS the mechanism for developing strong balance in bursty trains is different. There, in single events information can pass only one way and consecutive events in a train usually have reversed direction. Trains are possibly conversations between partners and the technology constrain is responsible for the strongly balanced communication. To simulate this behaviour again we generate trains as in the previous case by using Eq.6, however, to select the direction of the actual event we define a different mechanism. Here we assume that the direction of an event depends only on the direction of the previous one as they tend to be reversed. Accordingly, conditional probabilities can be used to decide the direction of the actual event as:

$$q_a(n|\sigma_{n-1}) = \frac{n}{n+1}, \quad q_b(n|\sigma_{n-1}) = 1 - \frac{n}{n+1} \quad (8)$$

where $\sigma \in \{A,B\}$ and $q_a(n|\sigma_{n-1})$ denotes the probability to choose the opposite direction for the $n$th event compared to the one in the $n-1$th step. Accordingly $q_b(n|\sigma_{n-1})$ denotes the probability of choosing the same direction as for the previous event. In this way the longer a train evolves, the larger is the probability to revert the direction of consecutive events and consequently the more the train becomes balanced.

The evolution of enhanced balance/inbalance in trains can be checked in the best way on edges where the overall communication is completely balanced and the balance/unbalance of trains is induced only by actual behavioural differences. In this way by setting $b_0 = 1/2$ the results obtained from the model process can be compared to averages calculated for real trains that evolve on edges with the same $b_0$ values. We selected edges with overall balance values between 0.5 $\leq b_0 < 0.55$ and calculate the corresponding $P(E)$ and $p(E)$ functions for this limited number of 115,277,534 calls and 69,288,504 SMS. As it is shown in Fig. 5b, the size of call trains (red circles) detected with $\Delta t = 600$s and SMS trains (blue squares) with $\Delta t = 300$s are distributed broadly and were characterized by an exponent $\beta = 4.6$ and $\beta = 3.5$ accordingly. Also the $p(E)$ balance values calculated for the limited event sets in Fig. 5c show similar behaviour to what was observed earlier in Fig. 4b for averages calculated for all edges. This figure demonstrates that even if the overall communication between people is balanced yet the strong communication unbalance for voice calls and an enhanced balance for SMS trains evolves. This is even more conspicuous when we compare these curves to the values obtained for independent processes (green triangles in Fig. 5c) using Eq.5 with $b_0 = 1/2$.

Using the above parameters we executed model processes for calls and SMS with the same number of events and corresponding $\gamma$ exponents deduced from $\beta$ according to Eq.6. The $P(E)$ distributions calculated for the evolving model trains are fitting well to the corresponding empirical functions as it is depicted in Fig. 5b for model call trains (black circles) and for model SMS sequences (yellow squares). It should be emphasized that there is no further fitting parameter in the model, nevertheless, the average $p(E)$ balance values calculated for trains in the model processes are in very good agreement with the empirical observations in Fig. 5c. This indicates that the assumed mechanisms are capturing rather accurately the salient features of the dynamics of directed human communication through phone calls and SMS. The only discrepancy is for the $p(E)$ values of short SMS trains, where the empirical data show an even-odd effect, which is not present in the model, indicating that for such communications an additional mechanism may be present enhancing the tendency towards the balance.

**Discussion**

In this paper we have studied mobile phone call and SMS communication sequences of millions of individuals and found the signature of internal correlations, which evolve between events and are responsible for long correlated event trains. By considering egocentric networks we realized that these bursty trains are likely to evolve between pairs of individuals rather than characterizing the communication of an individual in a larger group. Moreover, after a careful analysis we have found that the communication in such trains is much more balanced (unbalanced) for SMS (voice calls) than expected from trains of independent events with the same balance value $b_0$ of the edge, where the train was formed. We have concluded that the backgrounding mechanisms for the evolution of bursty trains can be interpreted as memory processes and implemented by reinforcement model dynamics. We have shown that events in trains are determined by different backgrounding motivations for calls and SMS; clearly in the latter case they are strongly mediated by technological constrains. We integrated these mechanisms into a model to simulate call and SMS train sequences. The model is capable in reproducing the observed statistics of events in trains of bursts and create correlations to enhance balance (imbalance) for SMS (voice call) cascades. We believe that this work gives insight into the communication dynamics of individuals, dyads, and egocentric networks at the level of single events. It points out similarities and differences of communication through different channels and contributes to our understanding of human behaviour.

**Materials and Methods**

In the present study we investigate the dynamics of human communication by analysing sequences of mobile-phone calls (MPC) and short messages (SMS) induced by a large set of individuals. The datasets were recorded by a single operator with 20% market share in an undisclosed European country (ethical statement was issued by the Northeastern University Institutional Review Board). They contain 633,986,311 time stamped MPC (209,316,760 SMS) events recorded during 182 days with 1 second resolution between 6,243,322 (4,819,993) individuals who are
connected via 16,783,865 (10,339,274) edges. In order to take into account only true social interactions and avoid commercial communication, we used only actions which were executed on links between users, who are at least once mutually connected each other during the recorded period. Note that due to the partial coverage we cannot know the complete egocentric network of each individual. This may play an effect on our results, however if this impact would be significant it would be already visible from the results of the limited dataset.

To characterize temporal behaviour of an individual, we commonly used inter-event times defined as
\[
\tau_i = \tau_{i+1} - (\tau_i + d_i)
\]
where \(\tau_i\) and \(\tau_{i+1}\) denote the starting time of two consecutive outgoing call events, while \(d_i\) is the corresponding duration. Naturally, in case of outgoing SMS the same definition holds but with \(d_i = 0\).

To filter the artificial effect of technology those consecutive short messages, which were sent to the same neighbour with inter-event times smaller than 10 seconds were replaced with a single event as they were possibly parts of the same multipart SMS [26]. Other technology and topology related constraints like answering machines for voice calls or degree heterogeneities in the network could be the possible sources for the emerging single link bursty train behaviour. To answer these questions we completed additional analysis with results discussed in the SI.

Supporting Information

Text S1 Properties of single link bursty behaviour. (PDF)

Acknowledgments

We thank M. Kivelä and J. Saramäki for comments and useful discussions and for A.-L. Barabási for the dataset used in this research.

Author Contributions

Conceived and designed the experiments: MK KK JK. Performed the experiments: MK. Analyzed the data: MK KK JK. Wrote the paper: MK KK JK. Constructed the model: MK.
References