ICTeCollective – Harnessing ICT-enabled collective social behaviour
Project no. 238597

Grant agreement: Small or medium-scale focused research project
Programme: FP7-ICT

Deliverable D5.2
[Report on model of within community dynamics]

Submission date: 2011-11-01

Start date of project: 2009-10-01 **Duration: 36 months**

Organisation name of lead contractor for this deliverable: BME

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	**X**

Document information

1.1 Author(s)

Author	Organisation	E-mail
J. Kertész	BME	kertesz@phy.bme.hu
A.-L. Barabási	BME, Northeastern Univ.	barabasi@gmail.com
K. Kaski	Aalto	kaski@gmail.com
J. Saramäki	Aalto	jari.saramaki@gmail.com
M. Karsai	Aalto	karsai.marton@gmail.com
R. Dunbar	Oxford	robin.dunbar@anthro.ox.ac.uk

1.2 Other contributors

Name	Organisation	E-mail
S. Vajna	BME	szabolcs.vajna@gmail.com
V. Palchikov	Aalto	palchykov@gmail.com
R. Pan	Aalto	rajkp@cc.hut.fi
H.-H. Jo	Aalto	h2jo23@gmail.com

1.3 Document history

Version#	Date	Change
V0.1	16.09.2011	Starting version, template
V0.2	30.09.2011	First version for circulation
V0.3	26.10.2011	Edited version
V1.0		Approved version to be submitted to EU

1.4 Document data

Keywords	Burstiness, temporal correlations, temporal motifs
Editor address data	kertesz@phy.bme.hu, karsai.marton@gmail.com
Delivery date	01/10/2011

1.5 Distribution list

Date	Issue	E-mail
01/11/2011	Consortium members	kaski@gmail.com
	Project officer	
	EC archive	

ICTeCollective Consortium

This document is part of a research project funded by the ICT Programme of the Commission of the European Communities as grant number ICT-2009-238597.

Aalto University (Coordinator)
School of Science
Department of Biomedical Engineering and Computational Science
FI-00076 AALTO, Espoo
Finland
Contact person: Prof. Kimmo Kaski
E-mail: kimmo.kaski@tkk.fi

Budapest University of Technology and Economics
Institute of Physics
Budapest, H-1111
Hungary
Contact person: Prof. János Kertész
E-mail: kertesz@phy.bme.hu

University of Oxford
Saïd Business School
The CABDyN Complexity Centre
Oxford, OX1 1HP
United Kingdom
Contact Person: Dr Felix Reed-Tsochas
E-mail: felix.reed-tsochas@sbs.ox.ac.uk

I.S.I Foundation
Torino 10133, Italy
Contact person: Dr. Santo Fortunato
E-mail: fortunato@isi.it

University of Warsaw
Faculty of Psychology
Warsaw 00927, Poland
Contact Person: Prof. Andrzej Nowak
E-mail: nowak@fau.edu

ICTeCollective introduction

ICTeCollective (*Harnessing ICT enabled collective social behaviour*) aims to develop systematic means of exploring, understanding and modelling2 systems where ICT is entangled with social structures. In particular, we will focus on behavioural patterns, dynamics and driving mechanisms of social structures whose interactions are ICT-mediated, from the level of individuals to the level of groups and large-scale social systems. Our unique approach is based on combined expertise in complex systems and the social sciences. By contrast with the majority of complexity studies that start from extremely simplified assumptions concerning social dynamics and concentrate on diagnosing structural features of social systems, we emphasize that ICT networks are dynamic systems of interacting humans and groups, and fully utilize the theories and methods of the social sciences are to be in ICTeCollective.

We will study and relate high quality datasets on ICT mediated social interactions and groups that have already been acquired, and also create new sets of data by conducting experiments with human subjects to examine the properties of social interactions mediated by technological means. The first source of data, electronic records of interactions, is a by-product of how ICT mediated communities operate. In particular, we will use some of the most extensive ICT datasets available at present, such as time-stamped data sets on mobile telephone communications between millions of users, the editing history of Wikipedia documents, and the popularity of Facebook applications. Secondly, entirely new data will be generated and released into the public domain by conducting laboratory experiments on ICT-mediated human interactions.
This project addresses the goals of the FP7 FET-OPEN call by trying to build an integrated picture of ICT-mediated social systems focussing on some aspects that are

i) critical to social interaction,
ii) can be easily tracked in large datasets and confirmed in experiments, and
iii) have a considerable

chance of improving our understanding and usage of ICT, with the possibility of leading to new and exciting technologies that can shape the future of ICT. The particular aspects that we focus on are *activity patterns, social influence*, and *group dynamics*. This choice helps us to address a large number of practical issues such as the driving mechanisms of social interactions mediated by ICT, and how these mechanisms then shape groups and society. All of these are critical to the goals of ICTeCollective.

We define the above terms as follows: *Activity patterns* are temporal sequences of social interaction and communication events, measurable in electronic communication records and representing the “atoms” of social interaction processes. *Social influence* refers to all processes where individuals affect each other’s beliefs, behaviour, activities, and representations of reality. *Group dynamics* comprises processes such as emergence, growth, merging, and splitting of groups, and associated behavioural patterns of individuals.

Executive Summary

The very inhomogeneous dynamic patter of human communication, which is partly responsible for the slow spreading in communication networks and runs under the notion of burstiness is usually characterized by an extremely broad distribution of inter-event times. However, it is difficult to capture the correlations behind such a dynamics. We have introduced a new measure of temporal behaviour of communicating individuals and proved that indeed the empirical patterns are highly correlated. We measured the number of events E within a well-defined bursty period and showed that the distribution $P(E)$ of that number is sensitive to the mentioned correlations. We could make a direct link between $P(E)$ and the memory of the process. Interestingly, very similar, correlated bursty time serried could be observed in a number of natural phenomena, showing the universal character of our findings. We implemented a simple two-state model to give a complete qualitative description of the empirical facts. The model contains a built in memory process and a reinforcement feedback mechanism. Using analytical tools we derived for a class of cases a scaling relationship between the exponents involved.

Understanding the patterns of human dynamics and social interaction, and the way they lead to the formation of an organized and functional society are important issues especially for techno-social development. Addressing these issues of social networks has recently become possible through large scale data analysis of e.g. mobile phone call records, which has revealed the existence of modular or community structure with many links between nodes of the same community and relatively few links between nodes of different communities. The weights of links, e.g. the number of calls between two users, and the network topology are found correlated such that intra-community links are stronger compared to the weak inter-community links. This is known as Granovetter's "The strength of weak ties" hypothesis. In addition to this inhomogeneous community structure, the temporal patterns of human dynamics turn out to be inhomogeneous or bursty, characterized by the heavy tailed distribution of inter-event time between two consecutive events. In this work, we study how the community structure and the bursty dynamics emerge together in an evolving weighted network model. The principal mechanisms behind these patterns are social interaction by cyclic closure, i.e. links to friends of friends and the focal closure, i.e. links to individuals sharing similar attributes or interests, and human dynamics by task handling process. These three mechanisms have been implemented as a network model with local attachment, global attachment, and priority-based queuing processes. By comprehensive numerical simulations we show that the interplay of these mechanisms leads to the emergence of heavy tailed inter-event time distribution and the evolution of Granovetter-type community structure. Moreover, the numerical results are found to be in qualitative agreement with empirical results from mobile phone call dataset

.

In most of the studies before the communication network as whole was investigated and considered as a proxy for the social interaction network. A traditional, complementary approach in sociology is to study the so-called egocentric networks. The new component from the point of view of computational social science is that we are now able to investigate a huge number of such egocentric networks, take averages over them and group the individuals according to their behaviour. Our first move in this direction was that we analysed the age and gender dependence of the first, second and third best “friends” (most frequently contacted persons) of individuals. In sociology it is widely accepted that homophily is a major friendship-forming factor. In most cases, however,

studies of homophily have focussed on psychological or social traits such as personality, interests, hobbies, religious or political views. Based on communication data we show homophily may also arise through a tendency for close friendships to be gender-biased. We also analyse the effect age on the observed pattern in egocentric networks.

Contents

1. Introduction

The data deluge due to ICT development has opened an entirely new approach in sociology. Instead of using surveys, case studies or observations on a rather limited set of target people, the sample of the size of whole society can be studied by investigating their "digital footprints". This data analysis, and the related modelling with massive computational demand is called Computational Sociology. Most of the efforts so far have been concentrated on static features like the topology of the network [1] and its relation to the intensity of the connections or to averaged dynamic properties, like mobility patterns [2] of individuals. The intrinsically inhomogeneous character of the human dynamics is usually taken into account by the broad distribution of the properties. Only recently it has become clear that this inhomogeneity is of major importance, e.g., in describing spreading phenomena in communication or contact networks [3,4]. The study of the bursty character [5] of individual dynamics, the dyadic patterns of communications has become a focus issue in this respect. The inhomogeneities are easily characterized by the distribution of inter-event times, however, this approach is unable to detect the *correlations* between the events. The traditional way of measuring correlations is hampered by the fat tailed inter-event time distribution. Therefore, there is a clear need to measure and model such correlations. Clearly, the first question in this context was to find the appropriate measure for the correlations. By defining the notion of *bursty periods* with respect to a time window, we were able to introduce a very sensitive quantity, the distribution $P(E)$ of the number of events in bursty periods. This is a plain exponential for *any* uncorrelated inter-event time distribution, thus deviations from an exponential clearly indicate non-trivial correlations. The correlations are related to memory effects and we have found an exact relationship between $P(E)$ and the description of the memory. The new measure of correlations has helped us to define a simple model, which is able to reflect qualitatively our empirical observations [6].

The observation of Granovetter-type community structure and individual bursty dynamics calls for integrating both structural and dynamical inhomogeneities into single framework or model in order to better understand the social dynamics with the smallest set of parameters. Although there are some approaches in integrating these structural and dynamical properties, the bursty nature of human behavior has been inherently assumed in these models. Instead, we are interested in the emergence of burstiness from the intuitive and natural model rules while at the same time generating the Granovetter-type community structure. In order to investigate the basic mechanisms responsible for various empirical observations, we incorporate the task handling process to the weighted network formation studied by Kumpula et al. [39]. In the presented model the weight assigned to a link is interpreted as the aggregate number of events on that link. Driven by both the cyclic and focal closure mechanisms a link is created by the first event occurring between individuals. Once created, the link is maintained by a series of events on that link, and finally removed by accidental memory loss of the individual. Each individual may initiate events or respond to those initiated by others, de- pending on the protocols determining the selection and execution of tasks given to individuals.

We assume that intensive communication means also close social ties. In humans, homophily (a striking tendency for individuals who share traits to preferentially form relationships) has emerged as an important organizing principle of social behavior [8]. In most such cases, studies of homophily have focussed on psychological or social traits such as personality, interests, hobbies, religious or political views. However, there is evidence that homophily may also arise through a tendency for close friendships to be

gender-biased [9]. Novel insights can be gained into both the structure of relationships and how the patterns of relationship change over the lifespan when the information that is available in large-scale mobile phone databases can be exploited in an appropriate way. We used such data to demonstrate striking differences in both mating and parental investment strategies between the two sexes.

These results are all pointing toward a better understanding of the structure of the society on small (inter community) scale. We are now able to zoom in into the network, characterize the individuals or their small groups by their dynamic behaviour, and identify similar groups and follow the mechanism of the formation of opinions.

2. Bursty correlations in human communication

2.1 Measuring intrinsic correlations with broad inter-event time distributions

Human communication is very inhomogeneous. It is characterized by bursts of activities, which are separated by long inactive periods [5]. This shows up in a communication network at the level of individuals and his/her partners (see Fig. 1). This bursty character is clearly recognizable in the digital records of human communication activities through different channels. During the last few years different explanations have been proposed to explain the origin of inhomogeneous human dynamics [5, 10, 11, 12]. The main discussion is about the impact of circadian and weekly fluctuations versus intrinsic correlations rooted in human nature and possibly related to decision-making processes or task handling. The latter was emphasized in a recent work using the novel technology of Radio Frequency ID's, where heterogeneous temporal behaviour explained by a reinforcement dynamics [13] driving the decision making process at the single entity level.

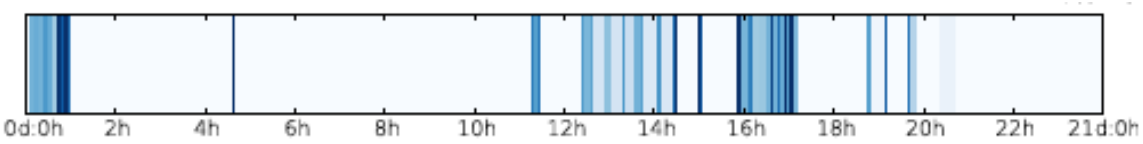


Figure 1. Bursty pattern of outgoing phone calls.

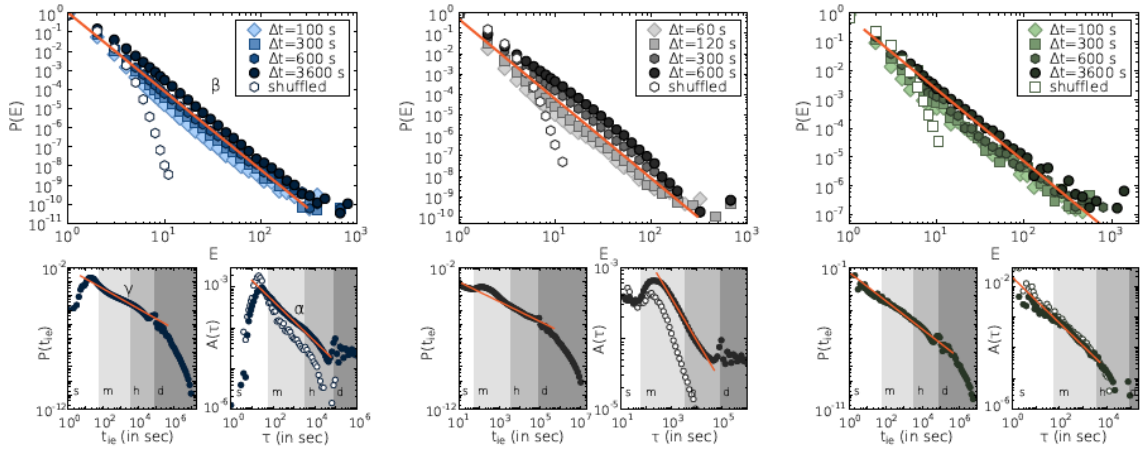
For systems with discrete event dynamics it is usual to characterize the observed temporal inhomogeneities by the inter-event time distributions, $P(t_{ie})$, where t_{ie} denotes the time between consecutive events. A broad $P(t_{ie})$, reflects large variability in the inter-event times and denotes heterogeneous temporal behaviour. Note that $P(t_{ie})$, alone tells nothing about the presence of correlations, characterized usually by the autocorrelation function, $A(\tau)$, or by the power spectrum density. However, for temporally heterogeneous signals of independent events with fat-tailed, power law $P(t_{ie})$, the Hurst exponent can assign spurious positive correlations together with the autocorrelation function [14]. To understand the mechanisms behind these phenomena, it is important to know whether there are true correlations in these systems. Hence for systems showing fat-tailed inter-event time distributions, there is a need to develop new measures that are sensitive to correlations but insensitive to fat tails.

A sequence of discrete temporal events can be interpreted as a time-dependent point process, $X(t)$, where $X(t) = 1$ at each time step, when an event takes place, otherwise it is zero. To detect bursty clusters in this binary event sequence we have to identify those events we consider correlated. The smallest temporal scale, at which correlations can emerge in the dynamics is between consecutive events. If only $X(t)$ is known, we can assume two consecutive actions to be related if they follow each other within a short time interval Δt. For events with the duration this condition is slightly modified.
This definition allows us to detect bursty periods, given as a sequence of events where each of them follows the previous one within a time interval Δt. By counting the number of events, E, that belong to the same bursty period, we can calculate their distribution $P(E)$ in a signal. For a sequence of independent events, $P(E)$ is uniquely determined by the inter-event time distribution $P(t_{ie})$ as follows:

$$P(E=n)=\left(\int_0^{\Delta t} P(t_{ie})dt_{ie}\right)^n\left(1-\int_0^{\Delta t} P(t_{ie})dt_{ie}\right)^2, \qquad (2.1)$$

which leads for finite Δt and for any $P(t_{ie})$ an exponential distribution for $P(E)$. Deviations from this exponential behaviour indicate correlations in the timing of the consecutive events.

To check the scaling behaviour of $P(E)$ in real communication systems we focused on outgoing events of individuals in three selected datasets: (a) A mobile-call dataset from a European operator; (b) Text message records from the same dataset; (c) Email communication sequences [15]. For each of these event sequences the distribution of inter-event times measured between outgoing events are shown in Fig.2 (left bottom panels) and the estimated power-law exponent values are summarized in Table 1. To explore the scaling behaviour of the autocorrelation function, we took the averages over 1; 000 randomly selected users with maximum time lag of $\tau = 10^6$. In Fig.2.a and b (right bottom panels) for mobile communication sequences strong temporal correlation can be observed (for exponents see Table 1). The power-law behavior in $A(\tau)$ appears after a short period denoting the reaction time through the corresponding channel and lasts up to 12 hours, capturing the natural rhythm of human activities. For emails in Fig.2.c (right bottom panels) long term correlation are detected up to 8 hours, which reflects a typical office hour rhythm (note that the dataset includes internal email communication of a university staff).



(a) Mobile-call sequence (b) Short message sequence (c) Email sequence

Figure 2. top: Distribution of the number of events in bursty periods. The inset shows the Δt-values. Bottom left: inter-event time distributions; bottom right: autocorrelation functions. a) Mobile-call data; b) SMS data; c) email data.

The broad shape of $P(t_{ie})$ and $A(\tau)$ functions confirm that human communication dynamics is inhomogeneous and displays non-trivial correlations up to finite time scales. . However, after destroying the correlations by shuffling the inter-event times in the sequences, the autocorrelation functions still show slow power-law like decay (empty symbols on bottom right panels), reflecting spurious positive correlations. This clearly demonstrates the disability of this function to characterize correlations for heterogeneous signals with fat tailed inter-event time distributions. However, for $P(E)$, we see a rather nice power law scaling with

$$P(E) \sim E^{-\beta}, \qquad (2.2)$$

for each of the event sequences as depicted in the main panels of Fig.2. Especially strong evidence for is that the roughly three orders of magnitude scaling is rather insensitive to changing Δt over a large interval, which corresponds to the scaling regime in the Figs. 2 b and c. However, after destroying the correlations by shuffling the inter-event times, we obtain an exponential $P(E)$, as expected. The effect is not due to the circadian or weekly patterns always present in human activities. Even by using a sophisticated de-seasoning technique [16] to get rid of these regularities, the remaining time series showed qualitatively the same results.

2.2 Relation to memory process

In each studied system qualitatively similar behavior was detected as single entities performed independent events or they passed through longer correlated bursty cascades. These cascades can be viewed as being the result of building up some stress in the system, e.g., the accumulation of important tasks in communication. Our hypothesis is that each bursty period is related to a single excitation that triggers a cascade relieving this stress, which would naturally explain the correlation between the events within the same train. Accordingly, we speculate that a single task execution can occasionally trigger a series of events responsible for the appearance of long bursty trains. This assumption is supported by the observation that bursty nodes communicate predominantly with only one of their neighbors [17], indicating that a bursty period maybe linked to a single task.
The correlations taking place between consecutive bursty events can be interpreted as a memory process, allowing us to calculate the probability that the entity will perform one more event within a Δt time frame after it executed n events previously in the actual cascade. This probability can be written as:

$$p(n) = \frac{P(E = n)}{\sum_{E=n}^{\infty} P(E)} \qquad (2.3)$$

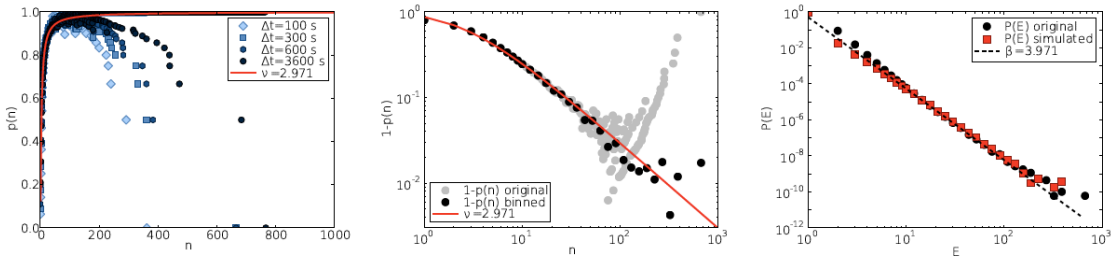
Therefore the memory function, $p(n)$, gives a different representation of the distribution $P(E)$. The $p(n)$, calculated for the mobile call sequence are shown in Fig.3a for trains detected with different window sizes. Note that in empirical sequences for trains with size smaller than the longest train, it is possible to have $p(n)$, = 1 since the corresponding probability would be $P(E = n) = 0$. At the same time, due to the finite size of the data sequence, the length of the longest bursty train is limited such that p(n) shows a finite cutoff.

We can use the memory function to simulate a sequence of correlated events. If the simulated sequence satisfies the scaling condition in (2.2) we can derive the corresponding memory function by substituting (2.2) into (2.3), leading to:

$$p(n) = \left(\frac{n}{n+1}\right)^{\nu}, \qquad (2.4)$$

with a scaling relation

$$\beta = \nu + 1. \qquad (2.5)$$



(a) Memory function of the mobile call sequence.

(b) Complementer memory function

(c) Empirical and simulated $P(E)$ distribution

Figure 3. Memory function (a), complementary memory function (b) and calculated $P(E)$ from the measured memory function and compared to the empirical curve.

In order to check whether (2.5) holds for real systems and whether the memory function in (2.4) describes correctly the memory in real processes we compare it to a memory function extracted from an empirical $P(E)$ distributions. We selected the $P(E)$ distribution of the mobile call dataset with Δt = 600 seconds and derived the corresponding $p(n)$ function. The complement of the memory function, in form of 1 - p(n), is presented in Fig.3.b where we show the original function with strong finite size effects (grey dots) and the same function after logarithmic binning (black dots). Taking equation (2.4) we _t the theoretical memory function to the log-binned empirical results using least-squares method with only one free parameter, ν. We find that the best fit offers an excellent agreement with the empirical data (see Fig.3.b and also Fig.3.a) with $\nu = 2.971 \pm 0.072$. This would indicate $\beta = 3.971$ through (5), close to the approximated value $\beta = 4.1$, obtained from directly fitting the empirical P(E) distributions in the main panel of Fig.2. In order to validate whether our approximation is correct we take the theoretical memory function p(n) of the form (2.4) with parameter $\nu = 2.971$ and generate bursty trains of 10^8 events. As shown in Fig.5.c, the scaling of the $P(E)$ distribution obtained for the simulated event trains is similar to the empirical function, demonstrating the validity of the chosen analytical form for the memory function.

2.3 Model study: Reinforcement dynamics with memory

We assume that the investigated systems can be described with a two-state model, where an entity can be in a normal state A, executing independent events with longer inter-event times, or in an excited state B, performing correlated events with higher frequency, corresponding to the observed bursts. To induce the inter-event times between the consecutive events we apply a reinforcement process based on the assumption that the longer the system waits after an event, the larger the probability that it will keep waiting. Such dynamics shows strongly heterogeneous temporal features as discussed in [13]. For our two-state model system we define a process, where the generation of the actual inter-event time depends on the current state of the system. The inter-event times are induced by the reinforcement functions that give the probability to wait one time unit longer after the system has waited already time t_{ie} since the last event. These functions are defined as

$$f_{A,B}(t_{ie}) = \left(\frac{t_{ie}}{t_{ie}+1} \right)^{\mu_{A,B}}, \qquad (2.6)$$

where μ_A and μ_A control the reinforcement dynamics in state A and B, respectively. These

functions follow the same form as the previously defined memory function in (2.4) and satisfy the corresponding scaling relation in (2.5). If $\mu_A << \mu_B$, the characteristic inter-event times at state A and B become fairly different, which induces further temporal inhomogeneities in the dynamics. The actual state of the system is determined by transition probabilities shown in Fig.4.b, where to introduce correlations between consecutive excited events performed in state B we utilize the memory function defined in equation (2.4).

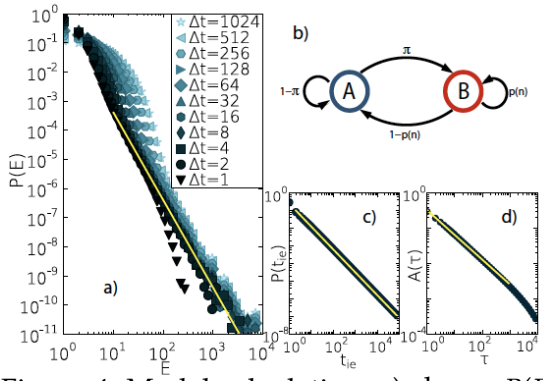


Figure 4: Model calculation. a) shows $P(E)$; b) the flow diagram of the model, c) and d) are the inter-event time distribution and the autocorrelation functions, respectively.

Figure 4 shows that the model is able to produce qualitatively satisfactory results, as all three characteristic functions have considerable scaling regimes, as observed in the empirical data.

2.4 Discussion

Using the measure we introduced, we were able to demonstrate for the first time the presence of intrinsic correlations in human communication patterns. Since we carefully analyzed the effect of circadian and other periodic patterns, as well as the effect of the heterogeneity of the users, we can conclude that a) Human communication is intrinsically bursty; b) This means not only a very broad, power law type inter-event time distribution but also long time memory related correlations. A simple phenomenological model was able to reflect the findings at least qualitatively. It remains to be understood, what influences the scaling exponents and how they are related to each other. Finally we mention that the correlated character of bursty time series was also demonstrated in natural phenomena, like earthquake activity patterns and neuron firing signals.

3. Emergence of bursts and communities in evolving weighted networks

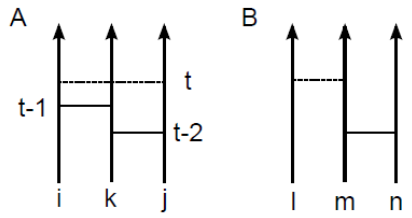
3.1 Motivation

Recently, large scale data analysis of mobile phone call records, has revealed the existence of modular or community structure with many links between nodes of the same community and relatively few links between nodes of different communities. The weights of links, i.e. the number of calls between two users, and the network topology are found correlated such that intra-community links are stronger compared to the weak inter-community links. This feature is known as Granovetter's "The strength of weak ties" hypothesis. In addition to this inhomogeneous community structure, the temporal patterns of human dynamics turn out to be inhomogeneous or bursty, characterized by the heavy tailed distribution of time interval between two consecutive events, i.e. inter-event time.

We study how the community structure and the bursty dynamics emerge together in a simple evolving weighted network model [26]. The principal mechanisms behind these patterns are social interaction by cyclic closure, i.e. links to friends of friends and the focal closure, i.e. links to individuals sharing similar attributes or interests, and human dynamics by task handling process. These three mechanisms have been implemented as a network model with local attachment, global attachment, and priority-based queuing processes. It is expected that the interplay of these mechanisms leads to the emergence of heavy tailed inter-event time distribution and the evolution of Granovetter-type community structure.

3.2 Model and results

We assume that the network evolves by means of link creation, link maintenance, and link deletion. Once a link between two stranger nodes is created by either the cyclic or the focal closure mechanisms, it is maintained by a series of events on that link, which we call the neighboring interaction (NI), or it is deleted by random memory loss. The focal closure mechanism is implemented by the random pairing of nodes, which is called global attachment or GA process. The cyclic closure mechanism is realized when a node interacts with its next nearest neighbor, which is called local attachment or LA process. While the GA process involves dyad interaction, the LA process is mediated by the third node, implying triad interaction. The NI process between neighboring nodes can happen directly, i.e. as dyad NI, or can be mediated by their common neighbor, i.e. as triad NI. Let us assume that only the event like the peer-to-peer phone call is considered. Then we can implement the triad interaction by splitting it into dyad interactions such that a node i has a chance to interact with j at time step t only when both i and j have interacted with the third node k recently, no more than, say, 2 time steps before (see Fig. 1).



We introduce the Triad-Interaction-enhanced model (TI model in short). The TI model is a direct extension of the weighted network model by Kumpula *et al.* [27], where the dyad NI process is analogous with Barabási's task execution model [5]. The dynamics at each time step t consists of the following three stages:

1) Triad Interaction (LA and triad NI): For each pair of nodes i and j satisfying $\{t_{ik},t_{jk}\}=\{t-2,t-1\}$ with a third node k, we check whether i and j are connected. If they are connected, an event between i and j occurs, i.e. $w_{ij} \rightarrow w_{ij} + 1$, corresponding to the triad NI process. Otherwise, the event between i and j occurred with probability p_{LA} leads to $w_{ij} = 1$, implying a link creation by the LA process. These LA and triad NI processes are responsible for the community formation and weight reinforcement, respectively.

2) Dyad Interaction (GA and dyad NI): Every node not involved in the previous stage selects a target node to make an event. If isolated, the node selects the target node from the whole population at random, preparing for the GA process. If non-isolated, the node selects the target node either from the whole population or from its neighbors with probabilities p_{GA} or $1 - p_{\mathrm{GA}}$, respectively. In other words, all nodes are free to find new neighbors while the non-isolated nodes are also responsible for maintaining links to the existing neighbors, the degree of which is controlled by p_{GA}. In the case of selecting the target from its neighbors, preparing for the dyad NI process, the probability of the node i selecting its neighboring node j is proportional to the weight between them, w_{ij}. Thus there is preference for the strong links. Targeting j by i is denoted by $i \rightarrow j$. One can make the analogy between the target selection from the population or from the neighbors and the task selection from the task list.

The nodes having selected their targets make events with targets in a random order only when both the node and its target are not yet involved in any other event at this time step. If the node i and its target j were not connected, the event would lead to a link creation between them, i.e. the realization of the GA process. Otherwise, the event between them results in $w_{ij} \rightarrow w_{ij} + 1$, implying the dyad NI process.

3) Memory Loss: With probability p_{ML}, each node, i, becomes isolated and a stranger to all its neighbors j as $w_{ij} = 0$. This completes the time step t.

Through all the above stages it has been assumed that the target has no choice to reject the event initiated by some other node. We term this the OR protocol [28] in a sense that it is enough for at least one of two nodes to initiate and make an event between them. Hence we call this version as the TI-OR model. Alternatively we can assume that an event can occur only in the reciprocal case, i.e. $i \rightarrow j$ and $j \rightarrow i$, which implies the AND protocol. It should be noted that for example a mobile phone user can reject a call from his/her friend by some reason. Here we will consider an TI-AND model, where the AND protocol is applied only to the dyad NI process.

As observables we obtain the cumulative weight distribution $P_c(w)$, the average number of next nearest neighbors $k_{nn}(k)$, the average overlap $O(w)$, the local clustering coefficient $c(k)$, the inter-event time distribution $P(\tau)$, and the average strength $s(k)$. In all cases results are averaged over 50 realizations for networks with $N = 50000$ and $p_{\mathrm{ML}} = 10^{-3}$. For the TI-OR model we obtain the average degree, $\langle k \rangle = 10.1$ and the global clustering coefficient, $\langle c \rangle = 0.08$ for $p_{\mathrm{LA}} = 0.013$ and $p_{\mathrm{GA}} = 0.1$. The cases with $p_{\mathrm{LA}} = 0.1$ and/or with $p_{\mathrm{GA}} = 0.07$ are plotted for comparison, see Fig. 2(left). For the TI-AND model we obtain $\langle k \rangle=9.6$ and $\langle c \rangle=0.13$ for $p_{\mathrm{LA}} = 0.07$ and $p_{\mathrm{GA}} = 0.1$. The cases with $p_{\mathrm{LA}} = 0.4$ and/or with $p_{\mathrm{GA}} = 0.04$ are plotted for comparison, see Fig. 2(right).

For the temporal dynamics the inter-event time distributions are characterized by the power-law with an exponential cutoff, i.e. $P(\tau) \sim \tau^{-\alpha} \exp(-\tau/\tau_c)$, where the scaling regimes span over about one decade. In case of TI-OR model, $\alpha = 2.5$ or 1.2 when $p_{LA} = 0.013$ or 0.1, respectively. In the case of TI-AND model, when $p_{LA} = 0.07$ or 0.4, we find $\alpha = 0.8$ or 0.6, respectively, both of which are close to the empirical value 0.7 of mobile phone call dataset within error bars.

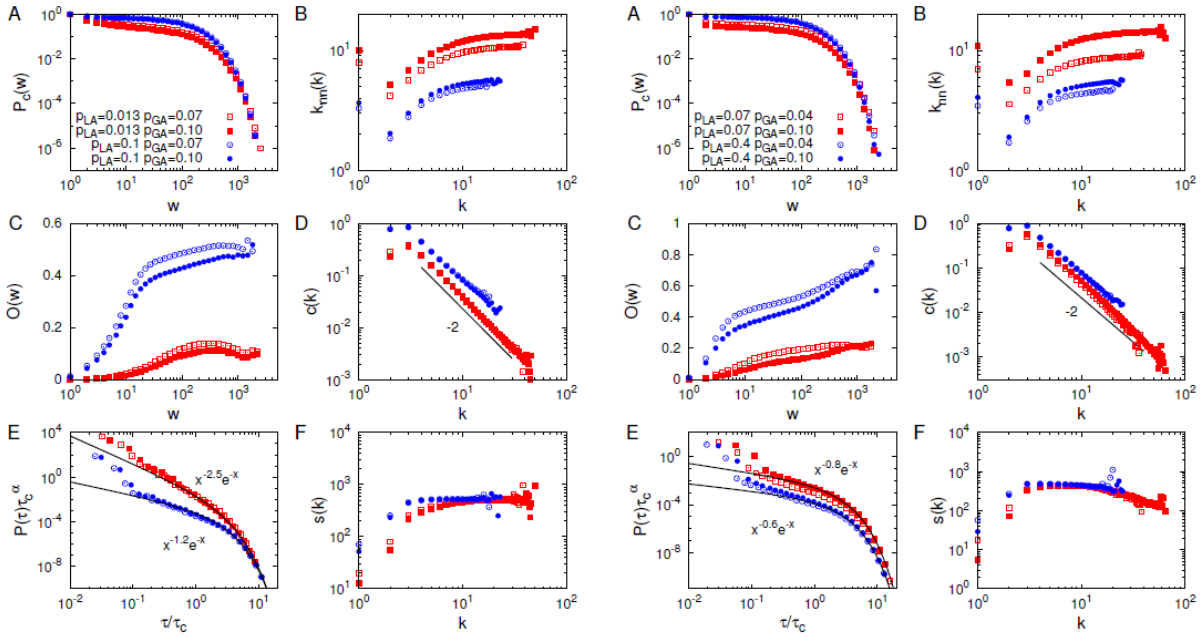


Figure 2. TI-OR model (left) and TI-AND model (right). See the text for details.

For both the TI-OR and TI-AND models, we find that the cumulative weight distributions are broad but do not follow power-law behavior for the various values of p_{LA} and p_{GA} as in the empirical analysis. It turns out that as the empirical results, the networks are assortative and have the Granovetter-type community structure, characterized by the increasing behavior of $k_{nn}(k)$ and $O(w)$, respectively. The sample networks shown in Fig. 3 also confirm the emergence of Granovetter-type community structure, such that the communities of internal strong links are connected by weak links. In addition, for the TI-OR model with $p_{LA} = 0.1$ we observe a slightly decreasing behavior of $O(w)$ for large w values, implying the existence of smaller but stronger communities. The decreasing behavior of the overlap was observed in the empirical analysis but not in the previous model studies [27].

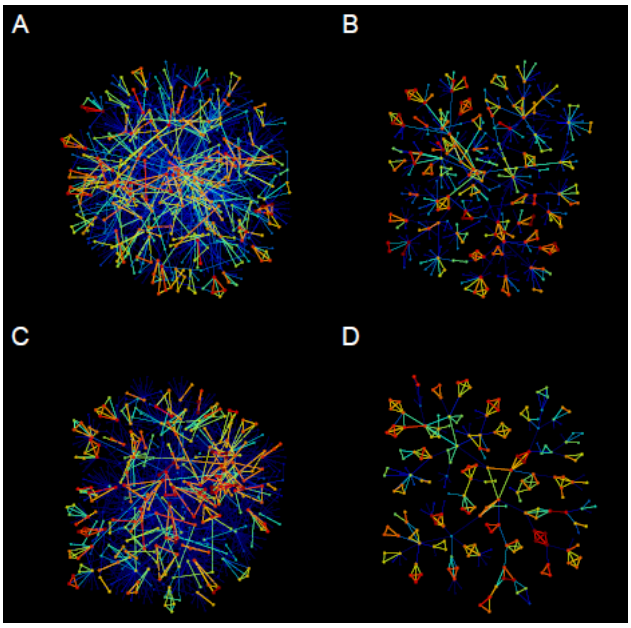


Figure 3. Snowball samples of networks for TI-OR (A, B) and TI-AND models (C, D). In B and D the links with weight 1 were removed for clear visualization. The color of links ranges from blue for weak links through yellow for intermediate links to red for strong links.

To figure out what are the possible underlying mechanism for these findings, we first identify the triangular chain interaction (TCI) among three neighboring nodes, say i, j, and k: Both the event between nodes i and j at time step t-2 and the event between nodes j and k at time t-1 lead to an event between nodes i and k at time t, again leading to another event between nodes i and j at time t+1 and so on, unless interrupted either by the events from/to nodes outside the triangle or by a random memory loss of nodes in the triangle. Since the TCI is exclusive due to the priority of the triad interaction including the LA process, the LA process enhanced by the large value of p_{LA} inhibits the interruption by the events from/to nodes outside the triangle, including the GA process, and thus making the community structure more compact in turn resulting in a smaller average degree. While the compact community structure enhances the TCI again, it can also make some neighbors of the TCI nodes wait for long time to interact with the TCI nodes. Hence, the larger value of p_{LA} gives rise to larger fluctuation in the inter-event times, implying the heavier tails of inter-event time distributions.

We have studied the emergence of Granovetter-type community structure, characterized by the increasing behavior of overlap as a function of the link weight, and the heavy tailed inter-event time distributions, i.e. bursty dynamics, in a single framework of simultaneously evolving weighted network model. By incorporating simple and intuitive task execution models for human dynamics into the weighted network model reproducing the Granovetter-type community structure of social systems, we successfully observe the qualitatively same behaviors as observed in the empirical networks based on the mobile phone call dataset. In addition, we have found that the exclusive triangular chain interaction (TCI) identified in the TI models plays the central role both in community structure formation and bursty dynamics.

4. Intimacy variations during life span

4.1 Egocentric networks and dyad formation

In sociology, the traditional way of studying the tie formation of an individual (ego) is the method of egocentric networks. The availability of metadata, like gender and age in the mobile call network enables us to use this method to study the gender and age dependence of dyadic connection with an unprecedented statistics.

In humans, homophily (a striking tendency for individuals who share traits to preferentially form relationships) has emerged as an important organizing principle of social behavior [18]. In most cases, however, studies of homophily have focussed on psychological or social traits such as personality, interests, hobbies, religious or political views. However, there is evidence that homophily may also arise through a tendency for close friendships to be gender-biased [19,20].

4.2 Data analysis and modeling

We used a very large mobile phone database to investigate gender preferences in close friendships: i) to test the hypothesis that preferences in the choice of the best friend are gender-biased; and ii) to investigate how these preferences change over the lifespan. We focus our attention on the three most preferred friends, as indexed by the frequency of contact. Several studies have demonstrated that frequency of contact is a reliable index of emotional closeness in relationships [21,22]. Recent research also reveals that personal social networks are hierachically structured [23,24], having a layer-like structure with distinct differences in the emotional closeness and frequency of contact with alters in the different layers, with an inner core of ~ 5 alters who between them account for about half our total social time.

In the dataset, all the calls as well as short text messages sent or received by each private subscriber was available over a seven month period. In order to filter out spurious effects like accidental or wrong number events as well as professional (e.g. call center) calls, we considered calls and text messages only between individuals that have at least one reciprocated contact (i.e. a contact in each direction). The original dataset contains information for about 6.8 million subscribers. In this study, we consider only those subscribers whose gender and age are both known. There are $N_M \approx 1.8$ million male and $N_F \approx 1.4$ million female subscribers for whom these data are available (see details below). To minimize errors due to individuals having multiple subscriptions, we took into account only those users with single subscriptions. Because both the user and the subscriber of the service (if different) have to be registered, we believe that the risk of double counting is insignificant.

For our analysis [25], we introduce a gender variable g, that can take the values $g_M = 1$ for males and $g_F = -1$ for females. Using the dataset, we define the average or representative gender of subscribers of known age and gender of users as follows:

$$\langle g \rangle = \frac{N_M}{N_M + N_F} g_M + \frac{N_F}{N_M + N_F} g_F,$$

which gives rise to an overall average gender bias for subscribers of $\langle g \rangle \approx 0.130$, indicating a general bias in favour of male alters. However, best friends with representative gender $\langle g_f \rangle \approx 0.007$, on average, are more female-biased than the

subscribers themselves, suggesting an overall preference for female alters as compared to random gender choice.

We first consider only those pairs of egos and alters whose age and gender information is available. This accounts for about 1.26 million ego/best-friend pairs, 0.84 million ego/second-best-friend pairs and 0.68 million ego/third-best-friend pairs.

In Fig. 8a, we present the average gender $\langle g_f \rangle^a$ of the best friend as a function of ego's age, for male and female egos separately. Until the age of ~50 years, both male and female egos prefer the best friend to be of opposite gender, although this effect is strongest for 30 years old males and 25 years old females.

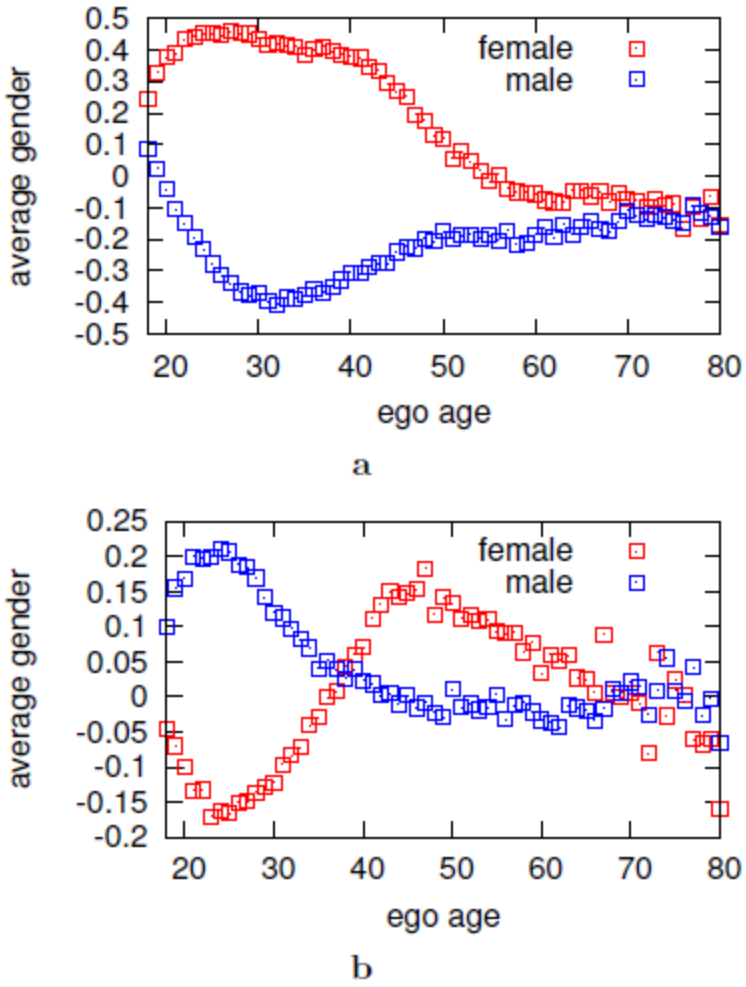


FIG. 8. a: The average gender of the best friend of an ego of specified age and gender. b: The average gender of the second best friend of an ego of specified age and gender. Blue squares correspond to male egos and red squares to female egos.

Notice that not only does the preference for an opposite-sex best friend kick in significantly earlier for females than for males (early 20s vs ~30 years of age: STATS), it

maintains a higher plateau for much longer. Whereas males exhibit a distinct and quite short-lived peak (a maximum of about 5 years from around age 30), women have a long, relatively higher male-biased plateau running from around the early 20s to as late as the early 40s – at which point, the male best friend seems to be moved into second place (Fig. 8b) and is replaced as best friend by a new (more typically female) alter. While males' best friends remain slightly female-biased throughout their lives, women's only eventually become so during their early 50s. The two sexes eventually converge on a slightly female-biased pattern at around 70 years of age.

The pattern for second best friends (Fig. 8b) is a partial mirror image of the pattern for best friends. Second best friends are typically same-sex, reaching a sharp peak in the early 20s before falling away gradually to reverse the gender bias in the late 30s. The transition is sharper for women than for men: males exhibit a shallower decline, and settle at a asymptotic value very close to gender equality, whereas women show a striking reversal to a strong male-biased peak in their late 40s (and a steady decline back towards equality by the late 60s). We ran a similar analysis for the gender of the third best friend: since the pattern is virtually identical to that in Fig. 8b. The similarity in the plots for second and third best friend reinforce the contrast to those for best friends, suggesting a more privileged status for the best friend.

In Fig. 9, we show age distributions of the best friends for male and female egos aged 25 and 50 years, respectively. On this finer scale of analysis, some additional patterns emerge. The distributions for friends of both genders are bimodal, with one maximum at around the ego's own age and the other with an approximately 25 year age difference. The maxima at ego's own age are opposite-gender biased, and most likely identify a male partner for female egos and vice versa for males. The maxima at the 25 year age difference (i.e. the generation gap) have a more balanced gender ratio, and most likely identifies, respectively, children and parents for 50-year-old egos and 25-year-old egos. The progression in this switch can be seen very clearly in the profiles for the intervening age cohorts.

These data allow us to draw three conclusions. First, women are much more focused on opposite-sex relationships than men are during the reproductively active period of their lives, suggesting that women invest more heavily in maintaining pair bonds than men do [23]. Second, as they age, women's attention shifts from their spouse to their children, but in particular to their daughters (reflected in the slight female-bias in older women's best friends). This transition is relatively smooth and slow for women (perhaps taking about 15 years to reach its new asymptote at around age 60), and may reflect the gradual arrival of grandchildren. Third, women in particular switch individuals around in their preference rankings much more than men do. Men tend to maintain a steadier pattern over a longer period, whereas women tend to switch individuals from one position to another in a more exaggerated way, perhaps reflecting shifts in their allegiances as their reproductive strategies switch more explicitly from mating to parental investment. This seems to be reflected in a tendency for women's gender-biases to be stronger than men's. Women's stronger investment in parental and grandparental investment is indicated by the fact that men's gender-biases for both best and second/third best friends show much

less evidence for any preference for contacting children: the younger (25-year) peak for 50-year men is half that for women's and shows a more even sex balance, whereas that for women is strongly biased in favour of female alters (i.e. daughters) (Fig. 9c,d).

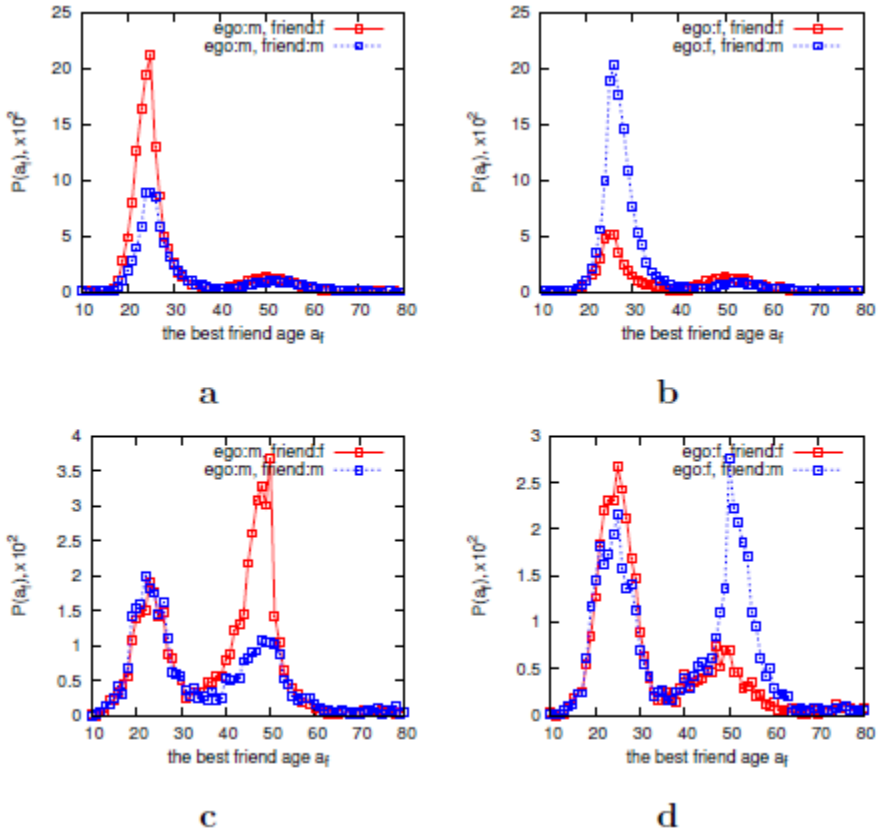


FIG. 9. The distributions P(af) of the best friend's age af for 25 years old male (**a**) and female (**b**) egos. In **c** and **d** we show the same distributions P(af) for 50 years old male and female egos, respectively. Red squares correspond to female best friends while blue squares to male best friends.

Fig. 10 plots age distributions of best friends for the two genders of ego. The profiles show clear differences demonstrating the well-known tendency of the age differences in romantic and marital relationships.

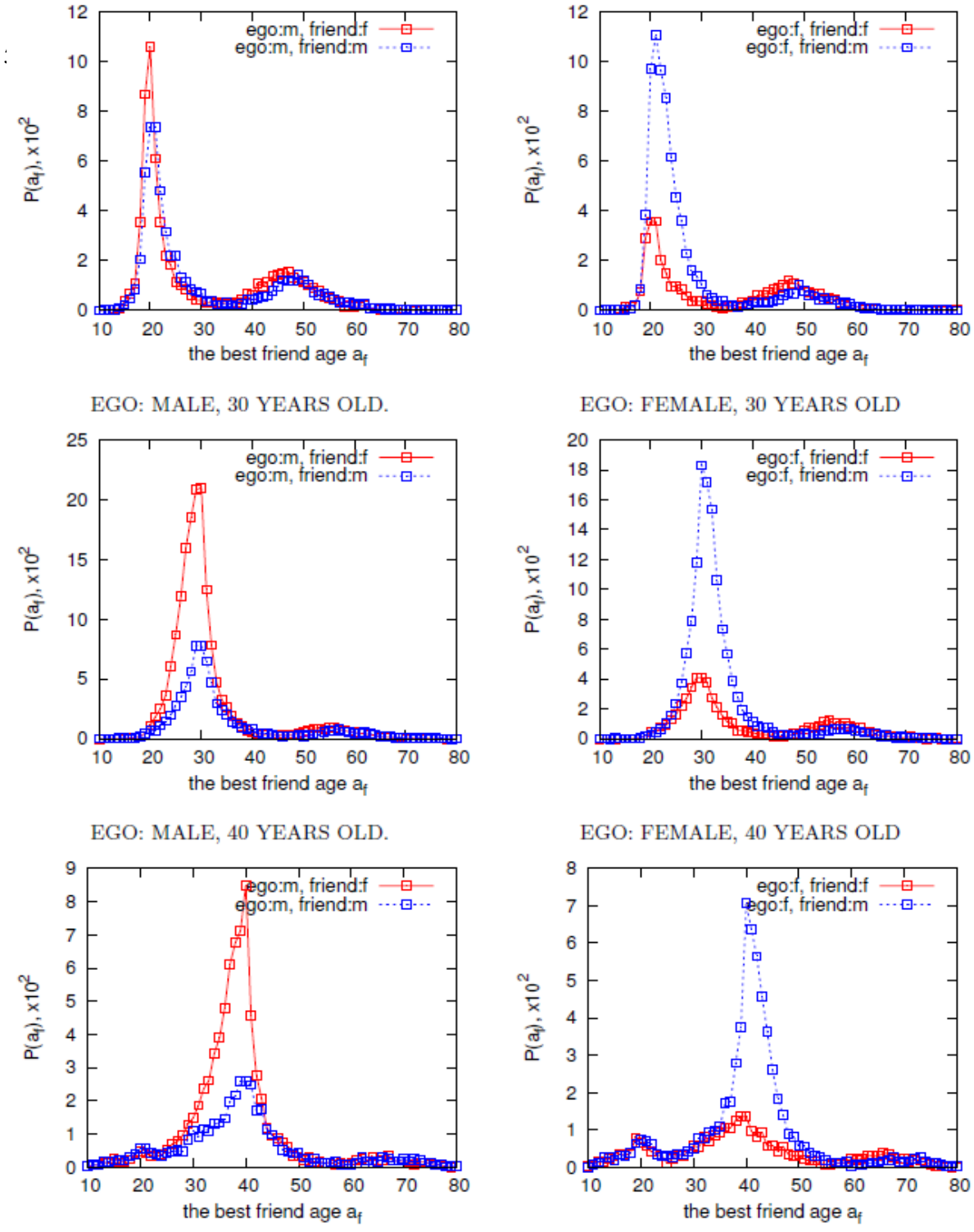


Figure 10. Gender and age dependence of best friends.

Furthermore, the analysis of the same kind of data for older age groups shows a clear double peaked structure, indicating that the parent-child relationships come into play. This tendency goes to the high ages, moreover, the peak with the age difference about 30 years becomes more and more dominant, indicating that elderly people have their close ties preferably with their adult children. Again, breaking down the data according to the

genders, reveals interesting differences between mother/daughter, mother/son, father/daughter and father/son relationships.

4.3 Discussion

While, inevitably, our analyses pool together large numbers of individuals, and so still lose some of the richness of the original data, nonetheless we have been able to demonstrate striking patterns in mobile phone usage data that reflect shifts in relationship preferences across the lifespan. Such patterns have not been noted previously, and open up novel opportunities for exploiting large network datasets of this kind. Aside from this purely methodological aspect, our analyses identify striking sex differences in the social and reproductive strategies of the two sexes that have not previously identified. One of the most important message of our work is that "digital footprints", like massive mobile phone billing data can be efficiently used to deep sociological studies.

References

[1] Onnela, J.-P., Saramäki, J., Hyvonen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L., Structure and tie strengths in mobile communication networks PNAS 104, 7332-7336 (2007)
[2] Gonzales, M., Hidalgo C., Barabási A.-L., Understanding individual human mobility patterns, Nature 453, 779 (2008)
[3] Karsai, M., Kivelä, M., Pan, R. K., Kaski, K., Kertész, J. Barabási, A.-L. , Saramäki, J., Small But Slow World: How Network Topology and Burstiness Slow Down Spreading, Phys. Rev. E **83**, 025102(R) (2011)
[4] Miritello, G., Moro, E., and Lara, R., Dynamical strength of social ties in information spreading, Phys. Rev. E 83 045102 (2011)
[5] Barabási, A.-L. The Origin of Bursts and Heavy Tails in Human Dynamics, Nature 435, 207 (2005)
[6] Karsai, M., Kaski, K., Barabási, A.-L., Kertész, J., Universal features of correlated bursty behavior, submitted
[7] Holme, P., Saramäki, J., Temporal Networks, arXiv:1108.1780
[8] Ip, G. W.-M., Chiu, C.-Y., & Wan, C. (2006). Birds of a feather and birds flocking together: physical versus behavioural cues may lead to trait- versus goal-based group perception. *Journal of Personality and Social Psychology* 90: 368-381.
[9] Roberts, S.B.G., Wilson, R., Fedurek, P. & Dunbar, R.I.M. (2008). Individual differences and personal social network size and structure. *Pers. Individ. Diffs.* 44: 954-964.
[10] Goh, K.-I. & Barab_asi, A.-L. Burstiness and memory in complex systems. Europhys. Lett. 81 48002 (2008).
[11] Malmgren, R.D. et al. A Poissonian explanation for heavy tails in e-mail communication. Proc. Natl. Acad. Sci. 105, 18153 (2008).
[12] Wu, Y. et al. Evidence for a bimodal distribution in human communication. Proc. Natl. Acad. Sci. 107, 18803 (2010).
[13] Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton ,J.F. & Vespignani, A. Dynamics of person-to-person interactions from distributed RFID sensor networks. PLoS ONE 5 e11596(2010)
[14] Vajna, S., Tóth, B., Kertész, J., Multi-agent processes with correlated burstiness (to be

published)
[15] Eckmann, J., Moses, E. & Sergi, D. Entropy of dialogues creates coherent structures in e-mail traffic. Proc. Natl. Acad. Sci. (USA) 101, 14333 (2004).
[16] Jo, H-H., Karsai, M., Kertész, J., Kaski, K., Circadian pattern and burstiness in human communication activity. arXiv:1101.0377 (2011)
[20] Karsai, M., et al. (to be published)
[21] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27: 415-444.
[22] Dunbar & R.I.M. and Spoors, M. Social networks, support cliques and kinship. *Human Nature* 6: 273-290.
[23] Roberts, S.B.G., Wilson, R., Fedurek, P. & Dunbar, R.I.M. (2008). Individual diff erences and personal social network size and structure. *Pers. Individ. Diffs.* 44: 954-964.
[24] Eagle, N., Pentland, A. & Lazerc, D. (2009). Inferring friendship network structure by using mobile phone data. *PNAS* 106: 1527415278.
[25] Roberts, S.B.G. & Dunbar, R.I.M. (2010). Communication in social networks: eff ects of kinship, network size and emotional closeness. *Pers. Relationships* 18: 439-452.
[26] Roberts, S.B.G., Dunbar, R.I.M., Pollet, T. & Kuppens, T. (2009). Exploring variations in active network size: constraints and ego characteristics. *Social Networks* 31: 138-146.
[27] R.A. Hill and R.I.M. Dunbar, Human Nature 14, 53-72 (2003).
[28] Palchykov, V., Kertész, J., Kaski, K., Barabási, A.-L., Dunbar, R.I.M., Sex differences in Intimacy Revealed by Mobile Phone Records Vary Across the Lifespan (to be published).
[38] H.-H. Jo, R. K. Pan, and K. Kaski, Emergence of Bursts and Communities in Evolving Weighted Networks, PLoS ONE 6, e22687 (2011)
[39] J. M. Kumpula, J. P. Onnela, J. Saramaki, K. Kaski, and J. Kertesz, Emergence of Communities in Weighted Networks, Physical Review Letters 99, 228701 (2007)
[40] B. Min, K. I. Goh, and I. M. Kim, Waiting time dynamics of priority-queue networks, Physical Review E 79, 056110 (2009)

OPEN ACCESS Freely available online

PLoS one

Emergence of Bursts and Communities in Evolving Weighted Networks

Hang-Hyun Jo*, Raj Kumar Pan, Kimmo Kaski

Department of Biomedical Engineering and Computational Science (BECS), Aalto University School of Science, Espoo, Finland

Abstract

Understanding the patterns of human dynamics and social interaction and the way they lead to the formation of an organized and functional society are important issues especially for techno-social development. Addressing these issues of social networks has recently become possible through large scale data analysis of mobile phone call records, which has revealed the existence of modular or community structure with many links between nodes of the same community and relatively few links between nodes of different communities. The weights of links, e.g., the number of calls between two users, and the network topology are found correlated such that intra-community links are stronger compared to the weak inter-community links. This feature is known as Granovetter's "The strength of weak ties" hypothesis. In addition to this inhomogeneous community structure, the temporal patterns of human dynamics turn out to be inhomogeneous or bursty, characterized by the heavy tailed distribution of time interval between two consecutive events, i.e., inter-event time. In this paper, we study how the community structure and the bursty dynamics emerge together in a simple evolving weighted network model. The principal mechanisms behind these patterns are social interaction by cyclic closure, i.e., links to friends of friends and the focal closure, links to individuals sharing similar attributes or interests, and human dynamics by task handling process. These three mechanisms have been implemented as a network model with local attachment, global attachment, and priority-based queuing processes. By comprehensive numerical simulations we show that the interplay of these mechanisms leads to the emergence of heavy tailed inter-event time distribution and the evolution of Granovetter-type community structure. Moreover, the numerical results are found to be in qualitative agreement with empirical analysis results from mobile phone call dataset.

Citation: Jo H-H, Pan RK, Kaski K (2011) Emergence of Bursts and Communities in Evolving Weighted Networks. PLoS ONE 6(8): e22687. doi:10.1371/journal.pone.0022687

Editor: Petter Holme, Umeå University, Sweden

Received June 11, 2011; **Accepted** June 28, 2011; **Published** August 12, 2011

Funding: Financial support by Aalto University postdoctoral program (Hang-Hyun Jo), by the Academy of Finland, the Finnish Center of Excellence program 2006–2011, project no. 129670 (Raj Kumar Pan, Kimmo Kaski) are gratefully acknowledged. Center of Excellence: http://www.aka.fi/en-GB/A/Centres-of-Excellence-/Ongoing/Centres-of-Excellence-in-Research-in-2006-2011/CoE-in-Computational-Complex-Systems-Research/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: joh1@aalto.fi

Introduction

Human dynamics and social interaction patterns have been a subject of intensive study in many different fields ranging from sociology and economics to computer science and statistical physics constituting what is nowadays called network science [1–4]. Partially due to the fact that huge amounts of various kinds of digital data on human dynamics have become available, explorative and quantitative analysis of these kinds of data has enabled us to have unprecedented insight into the structure and dynamics of behavioral, social, and even societal patterns. Examples of such data consist of email correspondence [5,6], mobile phone call (MPC) and Short Message (SM) communication [7–9], online social network services [10,11], and scientific collaboration [12].

The interaction structure among individuals in such large scale social data has been investigated by applying the concepts and methods of complex networks where individuals and their relationships represent nodes and links, respectively [13–15]. In many real networks, the link is characterized by a weight corresponding to the strength or closeness of social relationship [15,16], which in the case of MPC can be described by the aggregate number of calls between two individuals [7,17]. It has turned out that social networks are inhomogeneous and they can be characterized by modular or community structure [18]: The whole network is composed of separate communities connected by bridges, i.e. there are more and stronger links within communities than between communities, in accordance with Granovetter's "The strength of weak ties" hypothesis [19], corroborated later in [7,17]. This weight-topology coupling was successfully reproduced in the model of weighted networks driven by the cyclic and the focal closure processes [20]. Here the cyclic closure process refers to the link formation with one's next nearest neighbors, i.e. the link formation with friends of friends. The focal closure refers to the attribute-related link formation independently of the local connectivity [21]. It has been shown that these simple processes can lead to the emergence of complex weight-topology coupling, where the inhomogeneity of weights is a crucial factor for the emergence of communities.

In addition to the inhomogeneous community structure of social networks, the temporal patterns of human dynamics are inhomogeneous or bursty [6,22,23]. The bursts of rapidly occurring events of activity are separated by long periods of inactivity. The bursty dynamics is characterized by the heavy

tailed distribution of inter-event times τ, defined as the time interval between consecutive events, shows a power-law decay as $P(\tau) \sim \tau^{-\alpha}$ with $\alpha \approx 0.7$ or 1 for the MPC [23] or for the email [6], respectively. Two mechanisms for the origin of burstiness have been suggested: a) inhomogeneity due to the human circadian and weekly activity patterns [24,25] and b) inhomogeneity rooted in the human task execution [6,22]. Although such dynamic inhomogeneity is obviously affected by the circadian and weekly patterns, it was claimed that the burstiness turns out to be robust with respect to the removal of circadian and weekly patterns from the time series of MPC and SM activities [26]. Here we will concentrate on considering the dynamic inhomogeneities other than those due to circadian and weekly patterns, namely due to those related to individual behavior.

In relation to the inhomogeneity of human task execution, several priority-based queuing models have been studied [6,22,27–30]. Each individual is assumed to have a task list of finite size and select one of tasks under the selection protocol, such as selecting the task with the largest priority. Most of these models focus on the waiting time of task, which is defined as the time interval between the arrival time and the execution time of the task. However, in some cases, since the arrival times of tasks to the queue are not given, the waiting times cannot be empirically measured and thus cannot be directly compared to the empirical inter-event times. Furthermore, in spite of studying the communication patterns, such as the email correspondence, the interaction between individuals has not been properly considered in the models [6,22]. Some interactive models defined on networks assume that the underlying networks are binary and fixed [27–30]. However, in reality both the topology and the weights of social networks co-evolve according to the individual task executions as well as to social interaction by cyclic and focal closure mechanisms.

Both the structural inhomogeneity of social interaction and the dynamical inhomogeneity of human individual behavior affect the dynamical processes taking place on evolving social networks. For example it has been shown that the Granovetter-type weight-topology coupling slows down information spreading [7]. By using the analogy between link weight and information-bandwidth information turns out to spread fast and to get trapped within communities due to the internal strong links (broad bandwidth) and the weak links (narrow bandwidth) between communities, respectively. In addition to the effect of Granovetter-type community structure on information spreading individual bursty behavior also plays a crucial role in social dynamics. The long inactive periods represented by large inter-event times, inhibit the information spreading compared to the randomized null model, while the bursty periods of short inter-event times do not necessarily enhance the spreading [23]. Thus both the weight-topology coupling and the individual bursty dynamics should be taken into account and implement to a model in order to better understand the dynamics in the evolving social networks.

The observation of Granovetter-type community structure and individual bursty dynamics calls for integrating both structural and dynamical inhomogeneities into single framework or model in order to better understand the social dynamics with the smallest set of parameters. Although there are some approaches in integrating these structural and dynamical properties, the bursty nature of human behavior has been inherently assumed in these models [31,32]. Instead, we are interested in the emergence of burstiness from the intuitive and natural model rules while at the same time generating the Granovetter-type community structure. In order to investigate the basic mechanisms responsible for various empirical observations, we incorporate the task handling process to the weighted network formation studied by Kumpula *et al.* [20]. In our model the weight assigned to a link is interpreted as the aggregate number of events on that link. Driven by both the cyclic and focal closure mechanisms a link is created by the first event occurring between individuals. Once created, the link is maintained by a series of events on that link, and finally removed by accidental memory loss of the individual. Each individual may initiate events or respond to those initiated by others, depending on the protocols determining the selection and execution of tasks given to individuals.

Our model can be called co-evolutionary, in the sense that the task handling process of individuals affects the network evolution while the network structure constrains the individual behaviors. One of the typical issues in the co-evolutionary networks is that the timescale of network evolution competes with that of the dynamical process on the network [33,34]. In social dynamics the timescale for social relationship updates (a few weeks or months) is much larger than communication dynamics taking place on daily or hourly basis. In our case, since the events are the building blocks of the structure and the dynamics simultaneously, the relevant timescales are not explicitly controlled but emerged from the simple and intuitive rules of our model. In this paper, we show that by using the models with a few control parameters one can obtain the Granovetter-type community structure and also observe the emergence of bursty dynamics characterized by the heavy tailed inter-event time distribution.

This paper is organized such that we first introduce our two kinds of co-evolutionary models, the Triad-Interaction-enhanced model and the Process-Equalized model. Then we present the results for these models and discuss them in comparison with the empirical analysis results followed by the conclusions on the findings in the paper.

Methods

In our model we assume that the network evolves by means of link creation, link maintenance, and link deletion. Once a link between two stranger nodes is created by either the cyclic or the focal closure mechanisms, it is maintained by a series of events on that link, which we call the neighboring interaction (NI), or it is deleted by random memory loss. The focal closure mechanism is implemented by the random pairing of nodes, which is called global attachment or GA process. The cyclic closure mechanism is realized when a node interacts with its next nearest neighbor, which is called local attachment or LA process. While the GA process involves dyad interaction, the LA process is mediated by the third node, implying triad interaction. The NI process between neighboring nodes can happen directly, i.e. as dyad NI, or can be mediated by their common neighbor, i.e. as triad NI. Let us assume that only the event like the peer-to-peer phone call is considered. Then we can implement the triad interaction by splitting it into dyad interactions such that a node i has a chance to interact with j at time step t only when both i and j have interacted with the third node k recently, no more than, say, 2 time steps before, see fig. 1. In the following we propose two kinds of models. In the first kind the triad interaction takes place prior to the dyad interaction. We call this as Triad-Interaction-enhanced model (TI model in short). The TI model is a direct extension of the weighted network model by Kumpula *et al.* [20], where the dyad NI process is analogous with Barabási's task execution model [6]. In the second kind all the three processes (LA, GA, and NI) are considered equally and the framework of interacting and non-interacting tasks is adopted from [27], as the variant of Barabási's task execution model. We call this as Process-Equalized model (PE model in short).

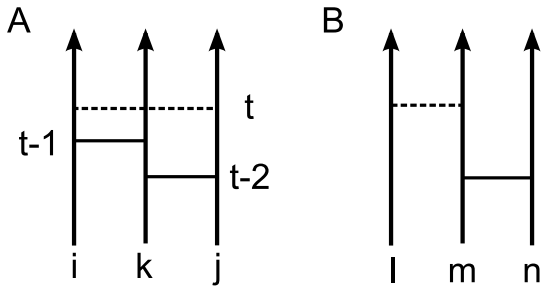


Figure 1. Schematic representation of local and global attachments of the model. Vertical lines represent the time lines of users. Horizontal solid and dashed lines represent events occurred and to be occurred between nodes, respectively. The number of events on a link defines the weight of the link. **A**. Local Attachment (LA): The node i has a chance to interact with node j at time step t only when there exists a temporal path connecting i and j through their common neighbor k within time window $[t-2,t-1]$. **B**. Global Attachment (GA): The isolated node l has a chance to interact with a randomly chosen node m.
doi:10.1371/journal.pone.0022687.g001

Now let us consider an undirected weighted network with N nodes. A weight of a link between nodes i and j, denoted by w_{ij}, can be interpreted as the aggregate number of events between them. The number of neighbors of node i is defined as the degree k_i. The time step of the most recent event between node i and node j is denoted by t_{ij}. Initially all nodes are set to be isolated, i.e. the initial network is without links.

Triad-Interaction-enhanced model

In the TI model the dynamics at each time step t consists of the following three stages:

1) Triad Interaction (LA and triad NI). For each pair of nodes i and j satisfying $\{t_{ik},t_{jk}\}=\{t-2,t-1\}$ with a third node k, we check whether i and j are connected. If they are connected, an event between i and j occurs, i.e. $w_{ij}\rightarrow w_{ij}+1$, corresponding to the triad NI process. Otherwise, the event between i and j occurred with probability p_{LA} leads to $w_{ij}=1$, implying a link creation by the LA process. These LA and triad NI processes are responsible for the community formation and weight reinforcement, respectively.

2) Dyad Interaction (GA and dyad NI). Every node not involved in the previous stage selects a target node to make an event. If isolated, the node selects the target node from the whole population at random, preparing for the GA process. If non-isolated, the node selects the target node either from the whole population or from its neighbors with probabilities p_{GA} or $1-p_{GA}$, respectively. In other words, all nodes are free to find new neighbors while the non-isolated nodes are also responsible for maintaining links to the existing neighbors, the degree of which is controlled by p_{GA}. In the case of selecting the target from its neighbors, preparing for the dyad NI process, the probability of the node i selecting its neighboring node j is proportional to the weight between them, w_{ij}. Thus there is preference for the strong links. Targeting j by i is denoted by $i\rightarrow j$. The analogy between the target selection from the population or from the neighbors and the task selection from the task list will be discussed later.

The nodes having selected their targets make events with targets in a random order only when both the node and its target are not yet involved in any other event at this time step. If the node i and its target j were not connected, the event leads to a link creation between them, i.e. the realization of the GA process. Otherwise, the event between them results in $w_{ij}\rightarrow w_{ij}+1$, implying the dyad NI process.

3) Memory Loss. With probability p_{ML}, each node, i, becomes isolated and a stranger to all its neighbors j as $w_{ij}=0$. This completes the time step t.

Through all the above stages it has been assumed that the target has no choice to reject the event initiated by some other node. We term this the OR protocol [28] in a sense that it is enough for at least one of two nodes to initiate and make an event between them. Hence we call this version as the TI-OR model. Alternatively we can assume that an event can occur only in the reciprocal case, i.e. $i\rightarrow j$ and $j\rightarrow i$, which implies the AND protocol. It should be noted that for example a mobile phone user can reject a call from his/her friend by some reason. Here we will consider an TI-AND model, where the AND protocol is applied only to the dyad NI process.

Process-Equalized model

In the TI model, since the triad interaction is executed prior to the dyad interaction, one can not control the intensity of the triad interaction. Therefore we have devised the PE model where we consider the triad interaction on the equal footing with the dyad interactions, i.e. the LA, GA, and NI processes are equally considered. In this case we incorporate the task execution process with interacting and non-interacting tasks [27], as described next.

Each node has the task list with one interacting task and one non-interacting task, denoted by I-task and O-task, respectively. The I-task represents the task requiring simultaneous interaction of two nodes, such as a phone call by a caller to a receiver, while the O-task represents some other task not requiring the simultaneity such as shopping, watching TV, etc. We count the inter-event times only for I-tasks, which settles down the issue of realistically interpreting the waiting time, as mentioned in [27]. The priorities of tasks are randomly drawn from the uniform distribution.

In this model the dynamics takes place such that at each time step t, every node selected in a random order goes through the stages 1) and 2). Then the stage 3) is performed:

1) Task and Target Selection. The node selects the task with larger priority. Only when it is I-task, this node, which we call a root node i, selects its target node either

- from the whole population with probability p_{GA}, i.e. the GA process, or
- from its next nearest neighbors with probability p_{LA}, i.e. the LA process, or
- from its neighbors with probability $1-p_{GA}-p_{LA}$, i.e. the NI process.

For the LA process, the next nearest neighbor of the root node is defined as the node j satisfying $\{t_{ik},t_{jk}\}=\{t-2,t-1\}$ with another node k. If the number of next nearest neighbors is more than 1, one of them is selected at random. For the NI process, the probability to target one of the root node's neighbors j is proportional to the weight w_{ij}, as in the TI model.

2) Task Execution. Only when the target node has not been involved in any event at this time, the event between the root node and the target node occurs, implying that the OR protocol is used. After this execution the priority of the I-task for the root node is replaced by the new random number while j's task list is not updated, implying that the target node did not execute its I-task but simply responded to the root node.

3) Memory Loss. Each node becomes isolated with probability p_{ML}, by which the time step t is completed.

Definitions of network properties

We calculate various network properties for the numerically obtained networks. Given the weight distribution $P(w)$, the cumulative weight distribution is defined by

$$P_c(w) \equiv \int_w^\infty P(w')dw'. \quad (1)$$

For each non-isolated node i, the number of next nearest neighbors, the individual clustering coefficient, and the strength are defined by

$$k_{nn,i} \equiv \frac{1}{k_i}\sum_{j\in\Lambda_i} k_j, \quad (2)$$

$$c_i \equiv \frac{2e_i}{k_i(k_i-1)}, \quad (3)$$

$$s_i \equiv \sum_{j\in\Lambda_i} w_{ij}, \quad (4)$$

respectively. Here Λ_i denotes the set of neighbors and e_i denotes the number of links among the node i's neighbors. The averages of the above quantities over the nodes with the same degree k define the average number of next nearest neighbors $k_{nn}(k)$, the local clustering coefficient $c(k)$, and the average strength $s(k)$, respectively. In addition, to test the Granovetter-type community structure the overlap is defined for each link connecting nodes i and j as follows:

$$O_{ij} \equiv \frac{|\Lambda_i \cap \Lambda_j|}{|\Lambda_i \cup \Lambda_j|}, \quad (5)$$

i.e. the fraction of the common neighbors over all neighbors of i and j. The average over the links with the same weight w defines the average overlap $O(w)$. For the dynamics we measure the inter-event time distributions $P(\tau)$.

Results and Discussion

The empirical analysis of mobile phone call data from a single operator in one European country for the first four months in 2007 [17,23] shows that $c(k)\sim k^{-\delta_c}$ with $\delta_c \approx 1$, $s(k)\sim k^{\delta_s}$ with $\delta_s \approx 1$. It also shows an increasing behavior of $k_{nn}(k)$, implying the assortativity, and an increasing behavior of $O(w)$ with slight decrease for very large w values, where the increasing part implies the Granovetter-type community structure. Moreover, it was found that $P(\tau)\sim\tau^{-\alpha}$ with $\alpha\approx 0.7$. In addition the average degree $\langle k\rangle$ turned out to be around 3.0 and the average clustering coefficient $\langle c\rangle$ around 0.13 when the new year's day of 2007 is excluded. It should be noted that the average degree of the mobile phone call network extracted from the single operator dataset might be underestimated compared to the full mobile phone call network composed by many operators. Therefore, we assume that the overall average degree of the whole social network is larger than 3, i.e. around 10. In this paper we consider the results to be relevant and comparable with reality, when $\langle k\rangle \approx 10$.

For the numerical simulations of the models described above we set the initial values as $N = 5\times 10^4$ and $p_{ML} = 10^{-3}$ for all the cases considered. The simulations of these models are found to become stationary at about $t = 3\times 10^3$, after which the numerical results are collected for 5×10^4 time steps.

Triad-Interaction-enhanced model

For both the TI-OR and TI-AND models, we find that the cumulative weight distributions are broad but do not follow power-law behavior for the various values of p_{LA} and p_{GA} as in the empirical analysis, see figs. 2 and 3. The similar behavior is found for degree and strength distributions (not shown). It turns out that as the empirical results, the networks are assortative and have the Granovetter-type community structure, characterized by the increasing behavior of $k_{nn}(k)$ for $k\geq 2$ and $O(w)$, respectively. Here most nodes with $k=1$ are supposed to be connected to randomly chosen nodes by the GA process, implying that $k_{nn}(1)\approx\langle k\rangle$. The sample networks shown in fig. 4 **A**–**D** also confirm the emergence of Granovetter-type community structure, such that the communities of internal strong links are connected by weak links. In addition, for the TI-OR model with $p_{LA}=0.1$ we observe a slightly decreasing behavior of $O(w)$ for large w values, implying the existence of smaller but stronger communities. The decreasing behavior of the overlap was observed in the empirical analysis but not in the previous model studies [20].

Based on the above observations it seems that a node is a member of a few strong triangles and connected to some other nodes outside its own triangles. This explains our finding of $\delta_c \approx 2$, different from the empirical result. It is because if the degree of a node increases mainly by means of the GA process, the number of links between neighbors remains while the number of all possible links grows as k^2, resulting in $c(k)\sim k^{-2}$. We also find $\delta_s \approx 0$ differently from the empirical value, which we will discuss later in relation to the dynamics.

In order to confirm the Granovetter-type community structure of networks, we perform the link percolation analysis. If links within communities are strong whereas links between them are weak as found in the empirical studies [7,17], the network should disintegrate faster when the weak links are removed first than when the strong links are removed first. Note that as shown in fig. 4 the links with weight 1 form an apparently random network as backgrounds for the community structure. Thus we apply the link percolation to the giant components of networks without links with $w=1$ and denote its size by N'. By removing links in an ascending or descending order of weights, we measure the remaining fraction of the giant component R_{GC}, the susceptibility χ, and the average clustering $\langle c\rangle$ as a function of the fraction of removed links, f. Here the susceptibility is defined as $\chi = \sum n_s s^2/N'$, where n_s denotes the number of clusters with size s and the giant component is excluded from the summation. For the weak-link-first-removal cases we find the sudden disintegration of networks at the finite value of f, i.e. $f_c = 0.62$ (0.21) for TI-OR (TI-AND) model. When the strong links are removed first, there is an apparent transition at the larger value of $f_c = 0.87$ (0.81) for TI-OR (TI-AND) model as shown in fig. 5 **A** and **C**. For the weak-link-first-removal cases the values of f maximizing $\langle c\rangle$, denoted by f_{max}, are quite close to those of f_c. When using the overlap instead of the weight for the link percolation, almost the same behavior is observed in fig. 5 **B** and **D** because $O(w)$ turns out to be the monotonically increasing function of w in our model.

For the temporal dynamics the inter-event time distributions are characterized by the power-law with an exponential cutoff, i.e. $P(\tau)\sim\tau^{-\alpha}\exp(-\tau/\tau_c)$, where the scaling regimes span over about one decade, see figs. 2 **E** and 3 **E**. In case of TI-OR model, $\alpha\approx 2.5$ or 1.2 when $p_{LA}=0.013$ or 0.1, respectively. In the case of TI-AND model, when $p_{LA}=0.07$ or 0.4, we find $\alpha\approx 0.8$ or 0.6,

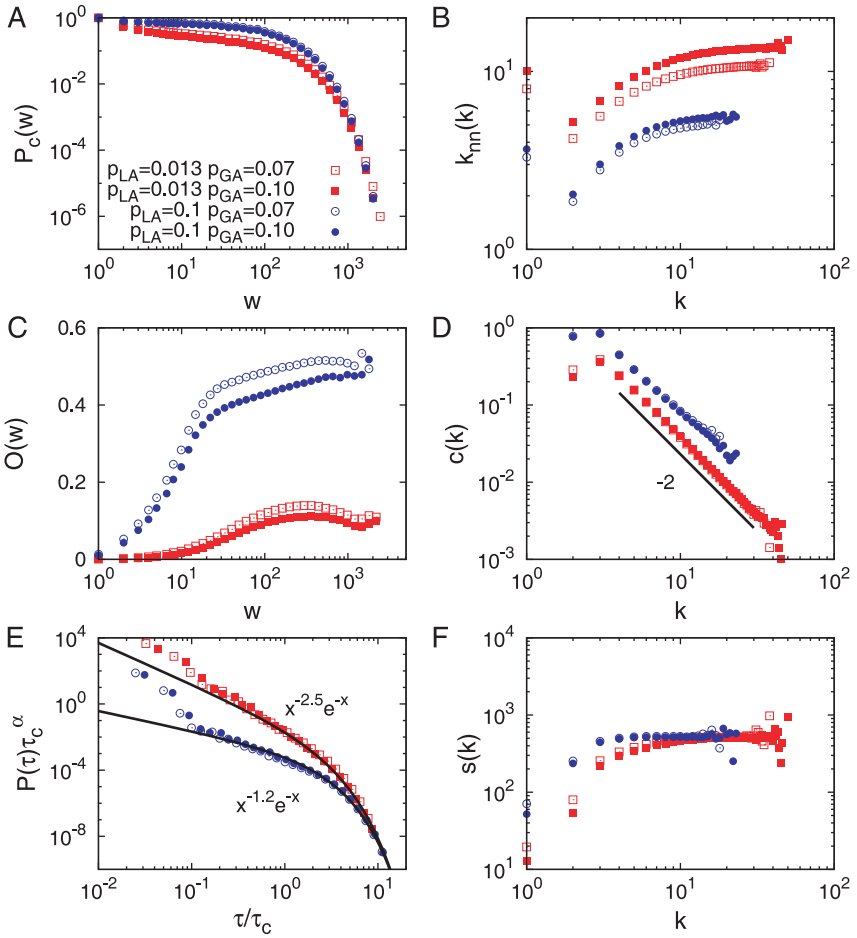


Figure 2. TI-OR model. A. The cumulative weight distribution $P_c(w)$. **B**. The average number of next nearest neighbors $k_{nn}(k)$. **C**. The average overlap $O(w)$. **D**. The local clustering coefficient $c(k)$. **E**. The inter-event time distribution $P(\tau)$. **F**. The average strength $s(k)$. Results are averaged over 50 realizations for networks with $N=5\times10^4$ and $p_{ML}=10^{-3}$. We obtain $\langle k\rangle\approx10.1$ and $\langle c\rangle\approx0.08$ for $p_{LA}=0.013$ and $p_{GA}=0.1$. The cases with $p_{LA}=0.1$ and/or with $p_{GA}=0.07$ are also plotted for comparison.
doi:10.1371/journal.pone.0022687.g002

respectively, both of which are close to the empirical value 0.7 of MPC dataset within error bars. In all cases, the values of α are smaller for larger values of p_{LA} but are barely affected by the value of p_{GA}. The values of τ_c turn out to be larger for larger values of p_{LA} and for smaller values of p_{GA}. The maximum value of τ_c is around 50.

To figure out what are the possible underlying mechanism for these findings, we first identify the triangular chain interaction (TCI) among three neighboring nodes, say i, j, and k: Both the event between nodes i and j at time step $t-2$ and the event between nodes j and k at time $t-1$ lead to an event between nodes i and k at time t, again leading to another event between nodes i and j at time $t+1$ and so on, unless interrupted either by the events from/to nodes outside the triangle or by a random memory loss of nodes in the triangle. Since the TCI is exclusive due to the priority of the triad interaction including the LA process, the LA process enhanced by the large value of p_{LA} inhibits the interruption by the events from/to nodes outside the triangle, including the GA process, and thus making the community structure more compact in turn resulting in a smaller average degree. In case of TI-OR model with $p_{GA}=0.1$, $\langle k\rangle\approx10.1$ or 4.2 for $p_{LA}=0.013$ or 0.1, respectively. While the compact community structure enhances the TCI again, explaining the observed peaks of $P(\tau)$ at $\tau=1$ and 2, it can also make some neighbors of the TCI nodes wait for long time to interact with the TCI nodes. Hence, the larger value of p_{LA} gives rise to larger fluctuation in the inter-event times, implying a smaller value of the power-law exponent α and a larger value of the cutoff τ_c, as observed. Based on this argument, the effect of p_{LA} dominates over that of p_{GA}, so that the value of p_{GA} barely affects the scaling of inter-event time distributions but it controls the value of τ_c. The larger value of p_{GA} allows nodes to choose a random target and thus interrupt the inter-event times of targets more frequently, leading to a smaller value of τ_c. The numerical results in the case of the TI-AND model can be explained by the same arguments, except for the observed values of α less than those found in the case of the TI-OR model. Note that in general the AND protocol inhibits the possibility of events.

The heavy tailed distribution of inter-event times, i.e. bursty dynamics, was not expected but it emerged from the model. Analogously with the task execution model suggested by Barabási [6], the dyad NI process can be interpreted such that a node i has the task list with size k_i and it selects one of neighbors (tasks) j with probability proportional to the priority of the task, i.e. the weight w_{ij} in our model. The degree k_i also varies depending on the link creation and deletion processes. A node having been isolated by

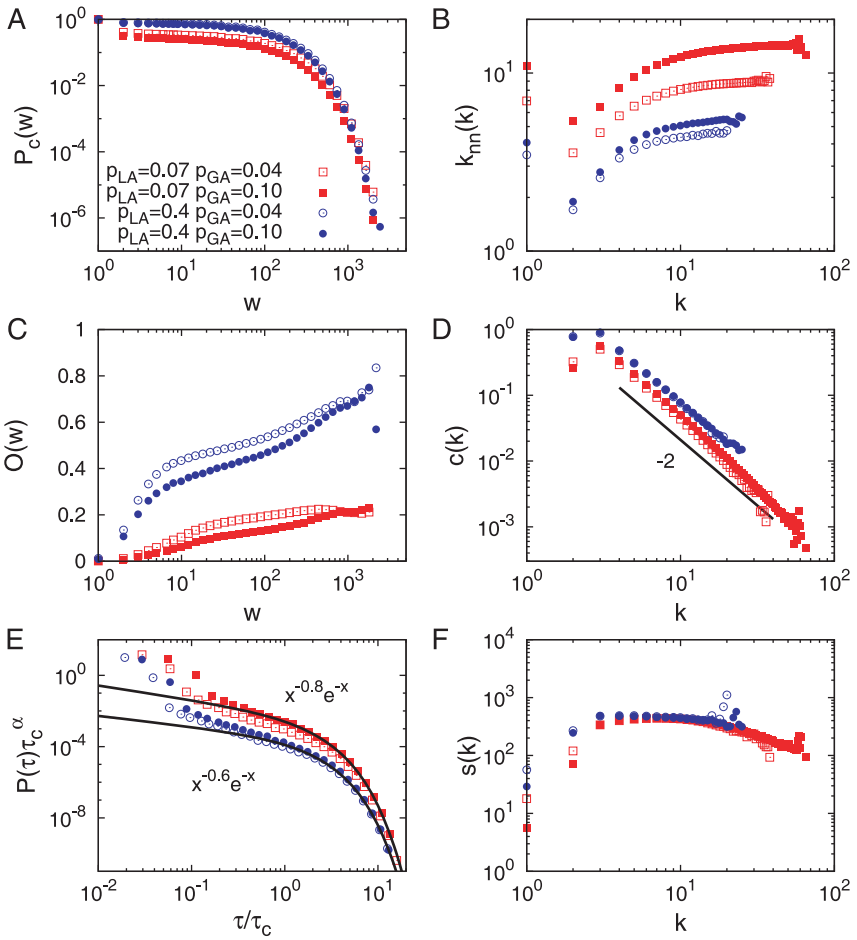


Figure 3. TI-AND model. A. The cumulative weight distribution $P_c(w)$. **B**. The average number of next nearest neighbors $k_{nn}(k)$. **C**. The average overlap $O(w)$. **D**. The local clustering coefficient $c(k)$. **E**. The inter-event time distribution $P(\tau)$. **F**. The average strength $s(k)$. Results are averaged over 50 realizations for networks with $N=5\times10^4$ and $p_{ML}=10^{-3}$. We obtain $\langle k\rangle\approx9.6$ and $\langle c\rangle\approx0.13$ for $p_{LA}=0.07$ and $p_{GA}=0.1$. The cases with $p_{LA}=0.4$ and/or with $p_{GA}=0.04$ are also plotted for comparison.
doi:10.1371/journal.pone.0022687.g003

the memory loss tries to interact with strangers. Once being connected to some other node by the GA process, its degree increases partly by means of the LA process but it will not diverge. The degree mostly fluctuates and sometimes remains unchanged for long periods of time. And the node finally becomes isolated again by the memory loss. Thus, the whole life-cycle of a node is assumed to consist of two types of periods, i.e. one with fixed-size and the other with variable-size task list. The periods of fixed-size task list, i.e. fixed degrees, are up to several hundred time steps, which are much larger than the observed τ_c. This implies the natural separation of timescales between network change and dynamics on the network, which is consistent with everyday experience of mobile phone usage. Due to the timescale separation the inter-event time distribution for the whole period can be represented by the superposition of those for fixed-size period and for variable-size period. Thus, to understand the effect of size variability on the scaling behavior of bursty dynamics, we refer to the previous works studied in the different kinds of models, such as by Vázquez *et al.* [22]. When the task list has a variable (fixed) size in the Barabási model, the power-law exponent for the waiting time distribution turns out to be 3/2 (2). According to the argument that the distribution of the inter-event times derived from the waiting times has the same power-law exponent as that of the waiting times, one can expect the similar values of exponent from our model. However, this is not the case with our model, so we leave this for the more rigorous analysis in the future.

Finally, the apparent overall independence of the average strength $s(k)$ on k for large values of k is attributed to the fact that once the node is a member of the TCI, its activity becomes effectively independent of its degree due to the exclusive property of TCI. We observe even the decreasing behaviors of $s(k)$ for the larger k values in the TI-AND model, i.e. the AND protocol based interaction with too many neighbors can make nodes failing to interact with any neighbors.

Process-Equalized model

The TI models show the expected behaviors of Granovetter-type community structure and the heavy tailed inter-event time distribution but they do not yield the expected behavior of the local clustering coefficient and average strength of the nodes. This is mainly due to too strong effect of the triad interaction and that is why we need to consider the PE model for modeling improvement and comparison with empirical results.

With the PE model we find that the cumulative weight distributions $P_c(w)$ are broad, that the overlap $O(w)$ increases with w, i.e. showing Granovetter-type community structure, that

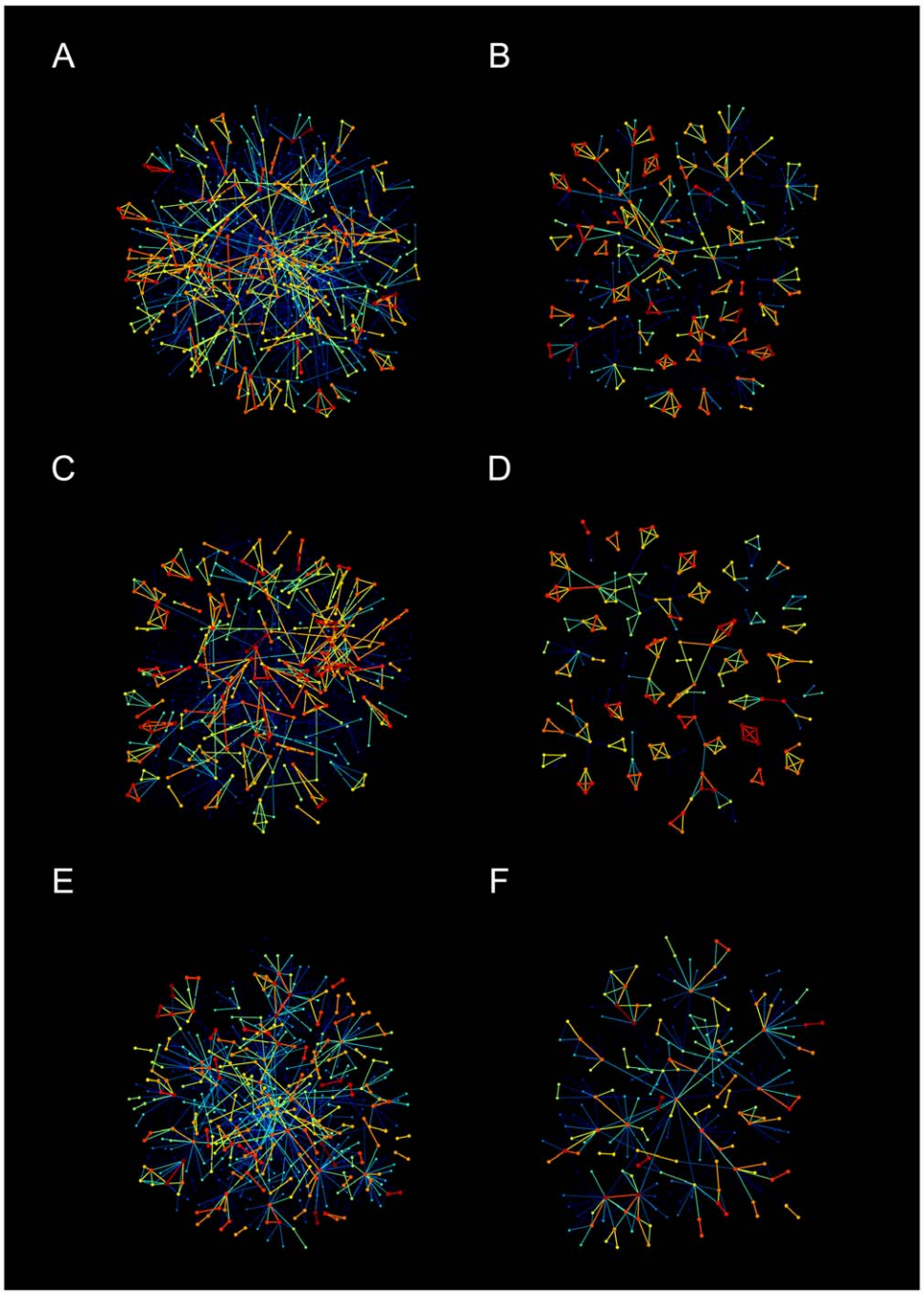


Figure 4. Snowball samples of networks [**36**]**. A**, **B**. TI-OR model. **C**, **D**. TI-AND model. **E**, **F**. PE model. For each model, we plot the sample network starting from a random node (left panel) and the one without the links with $w=1$ (right panel) for clear visualization. The color of links ranges from blue for weak links through yellow for intermediate links to red for strong links.
doi:10.1371/journal.pone.0022687.g004

the average number of next nearest neighbor $k_{nn}(k)$ increases with k, i.e. showing the network being assortativity for $k \geq 2$, and that $c(k) \sim k^{-\delta_c}$ with $\delta_c \approx 1$ and $s(k) \sim k^{\delta_s}$ with $\delta_s \approx 1$, as shown in fig. 6. All these results are consistent with the empirical analysis on real data. Based on the sample networks in fig. 4 **E** and **F**, it is evident that the TCI becomes weaker and less exclusive than in the case of the above TI models. Therefore, as the degree of a node increases, the neighbors of that node have the increasing chance to interact with each other, resulting in $c(k) \sim k^{-1}$.

In fig. 5 **E** and **F** we show the results of the link percolation analysis, done to confirm Granovetter-type community structure. We find that when the weak links are removed first, the percolation transition occurs at $f_c = 0.86$. On the other hand when the strong links are removed first, a transition is observed at $f_c = 0.91$, implying that the strong links play the role of bridges between communities. This is also evident in the sample networks in fig. 4 **E** and **F**. The curve of the average clustering coefficient $\langle c \rangle$ turns out to be flat for a wide range of f values. Similar

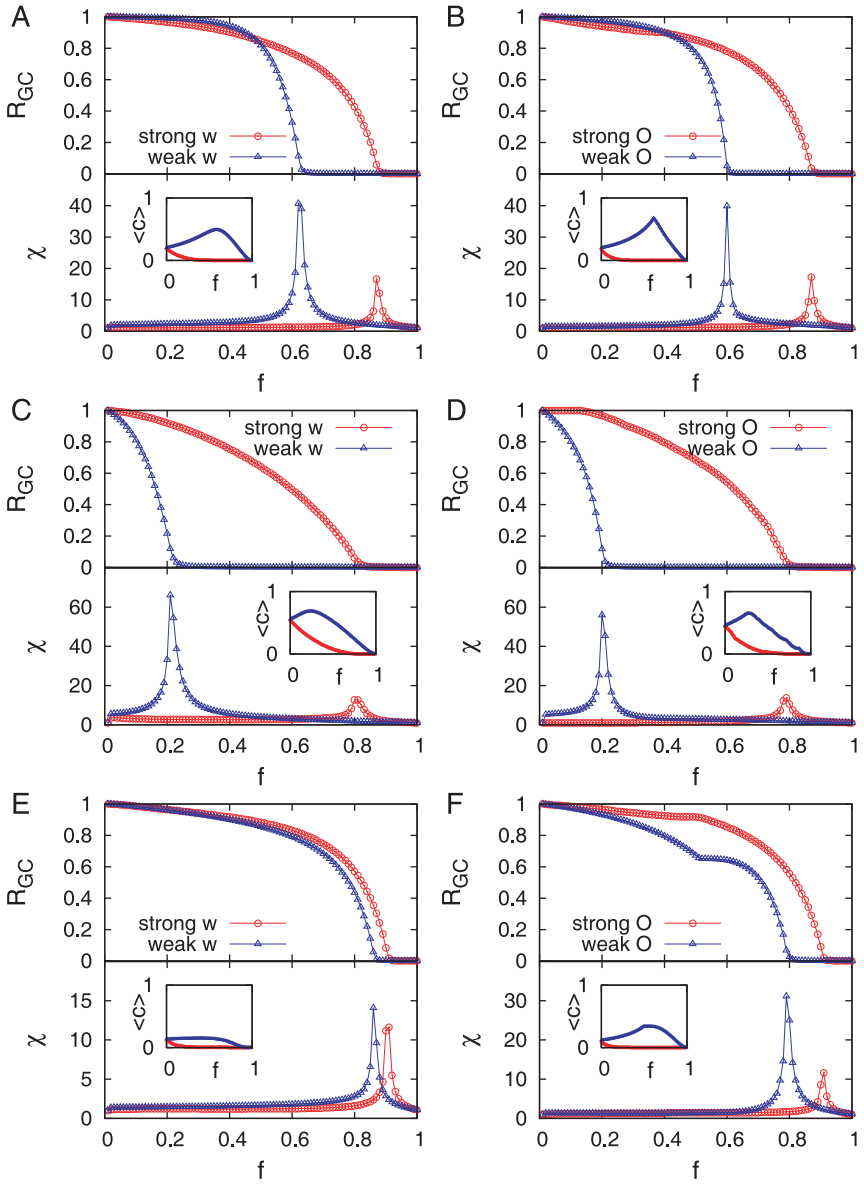


Figure 5. Link percolation analysis. A, **B**. TI-OR model. **C**, **D**. TI-AND model. **E**, **F**. PE model. As the link strength, we use the weight (left panel) and the overlap (right panel). For each panel, we calculate the fraction of giant component R_{GC}, susceptibility χ, and clustering coefficient $\langle c \rangle$ (inset) as a function of the fraction of removed links, f. Results are averaged over 50 realizations for networks originally with $\langle k \rangle \approx 10$ for each model.
doi:10.1371/journal.pone.0022687.g005

behaviors are also observed when the overlap is used instead of the weight in the link percolation analysis. For the weak-link-first-removal we find $f_c = 0.79$ and $f_{max} = 0.55$, where yet another kink in the curve of R_{GC} is observed. This implies that the network goes through two abrupt changes, first at f_{max} and then at f_c.

Here we also observe the heavy tailed distributions of inter-event times with exponential cutoffs following a power law behavior with the exponent of $\alpha \approx 1.1$. The task execution model for each node would result in $\alpha = 1$ as in the case of Barabási's queuing model if only initiating the I-tasks are counted as the relevant events and if the neighbors of the node always respond to that node. However, the nodes are supposed to interact with each other such that by initiating I-tasks some root nodes can interrupt the inactive periods of their target nodes, which in general decreases the inter-event times. On the other hand, if the target is already involved in another event so that the trials by the root nodes fail, the inter-event times of corresponding root nodes would increase up to the points of next successful events occurring. The observed value of $\alpha \approx 1.1$ indicates that any of the mentioned factors did not affect much the scaling behavior of the distributions. The values of τ_c are largely or barely affected by the value of p_{GA} or p_{LA}, respectively, in an anti-correlated way. The maximum value of τ_c is around 270.

The observation of the average strength $s(k) \sim k$ behavior can be explained by considering the dynamics where the OR protocol is adopted. In this case the nodes with many neighbors might receive more calls from their neighbors than those with few neighbors do, while the chance to make calls is the same for any node.

Conclusions

We have studied the emergence of Granovetter-type community structure, characterized by the increasing behavior of overlap

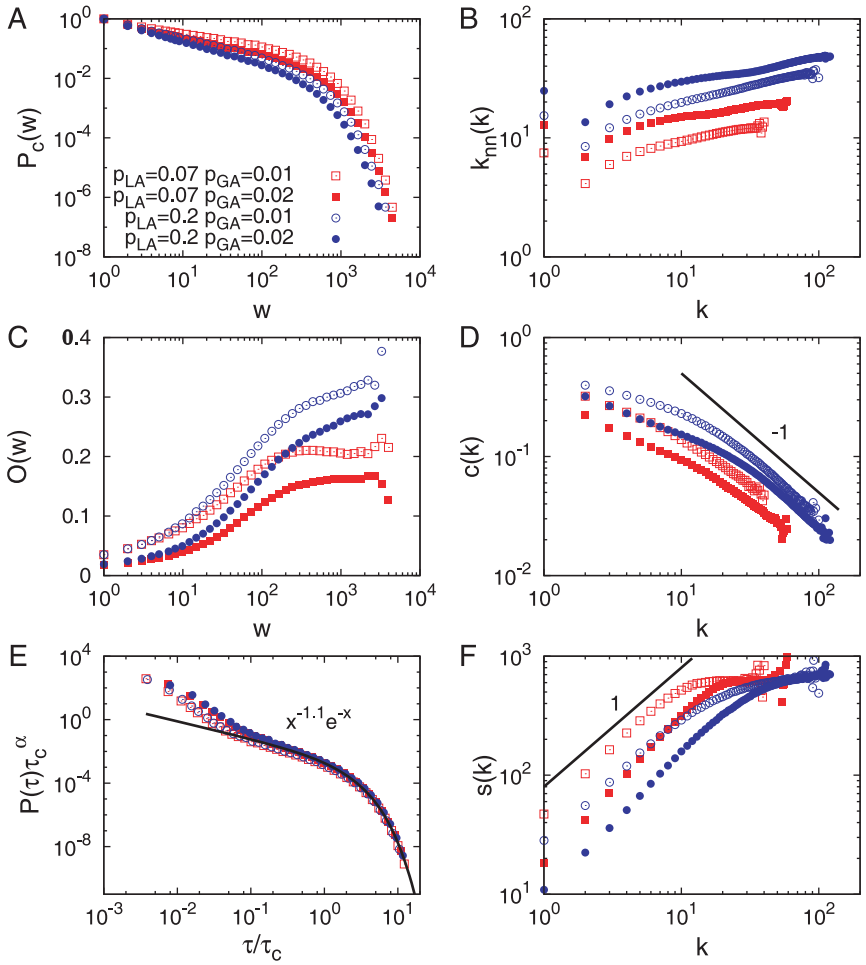


Figure 6. PE model. A. The cumulative weight distribution $P_c(w)$. **B**. The average number of next nearest neighbors $k_{nn}(k)$. **C**. The average overlap $O(w)$. **D**. The local clustering coefficient $c(k)$. **E**. The inter-event time distribution $P(\tau)$. **F**. The average strength $s(k)$. Results are averaged over 50 realizations for networks with $N=5\times10^4$ and $p_{ML}=10^{-3}$. We obtain $\langle k\rangle\approx9.9$ and $\langle c\rangle\approx0.11$ for $p_{LA}=0.07$ and $p_{GA}=0.02$. The cases with $p_{LA}=0.2$ and/or with $p_{GA}=0.01$ are also plotted for comparison.
doi:10.1371/journal.pone.0022687.g006

as a function of the link weight, and the heavy tailed inter-event time distributions, i.e. bursty dynamics, in a single framework of simultaneously evolving weighted network model. By incorporating simple and intuitive task execution models for human dynamics into the weighted network model reproducing the Granovetter-type community structure of social systems, we successfully observe the qualitatively same behaviors as observed in the empirical networks based on the mobile phone call (MPC) dataset. In addition, we have found that the exclusive triangular chain interaction (TCI) identified in the TI models plays the central role both in community structure formation and bursty dynamics. For the existence of TCI we have the evidence from the empirical study on the dynamic motifs of MPC communication [35]. The numerical results from TI-OR and TI-AND models are qualitatively the same except for the power-law exponent α of inter-event time distributions. The values of α from TI-AND model turn out to be closer to the empirical value 0.7 for the MPC, implying that the AND protocol is necessary to properly model the MPC communication. Furthermore, in the PE model, by relaxing the exclusive property of TCI to some extent we could obtain more realistic results at least for the network structure. The scaling behavior of inter-event time distributions seems to be mainly affected by the incorporated framework of interacting and non-interacting tasks, which should be made clear in the future.

Finally we believe that building simple empirical-observation-based models, like our TI- and PE-models, by incorporating the process of human task execution by priority-based queuing with the basic processes of friendship-network formation by cyclic and focal closure mechanisms enable us to better understand the underlying mechanisms of real co-evolutionary networks. Furthermore, these models enable us to explore the social dynamics in these networks as done differently by Karsai *et al.* [23] with the susceptible-infected (SI) dynamics for the mobile phone call communication. Moreover, the scaling properties and finite-size scaling of real networks are usually not so informative but can be considered and made more informative by means of simple but still quite realistic models, where one can control the system size and other parameters as well.

Author Contributions

Conceived and designed the experiments: HJ RKP. Performed the experiments: HJ RKP. Analyzed the data: HJ RKP. Wrote the paper: HJ RKP KK.

References

1. Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences). Structural analysis in the social sciences, 8. Cambridge University Press, 1 edition.
2. Goyal S (2007) Connections : An introduction to the economics of networks. Princeton, NJ: Princeton University Press.
3. Newman MEJ (2010) Networks: An Introduction. USA: Oxford University Press, 1 edition.
4. Barabási AL (2010) Bursts: The Hidden Pattern Behind Everything We Do. Dutton Books.
5. Eckmann JP, Moses E, Sergi D (2004) Entropy of dialogues creates coherent structures in e-mail trafic. Proceedings of the National Academy of Sciences of the United States of America 101: 14333–14337.
6. Barabási AL (2005) The origin of bursts and heavy tails in human dynamics. Nature 435: 207–211.
7. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, et al. (2007) Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences 104: 7332–7336.
8. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. Nature 453: 779–782.
9. Wu Y, Zhou C, Xiao J, Kurths J, Schellnhuber HJ (2010) Evidence for a bimodal distribution in human communication. Proceedings of the National Academy of Sciences 107: 18803–18808.
10. Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In:Proceedings of the 19th international conference on World wide web New York, NY, USA: ACM, WWW 10: 591–600. doi:10.1145/1772690.1772751.
11. Onnela JP, Reed-Tsochas F (2010) Spontaneous emergence of social inuence in online systems. Proceedings of the National Academy of Sciences 107: 18375–18380.
12. Newman MEJ (2001) The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the United States of America 98: 404–409.
13. Albert R, Barabasi AL (2002) Staötistical mechanics of complex networks. Reviews of Modern Physics 74: 47–97.
14. Newman MEJ (2003) The structure and function of complex networks. SIAM Review 45: 167–256.
15. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D (2006) Complex networks: Structure and dynamics. Physics Reports 424: 175–308.
16. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America 101: 3747–3752.
17. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Menezes, et al. (2007) Analysis of a large-scale weighted network of one-to-one human communication. New Journal of Physics 9: 179.
18. Fortunato S (2010) Community detection in graphs. Physics Reports 486: 75–174.
19. Granovetter MS (1973) The strength of weak ties. The American Journal of Sociology 78: 1360–1380.
20. Kumpula JM, Onnela JP, Saramäki J, Kaski K, Kertész J (2007) Emergence of communities in weighted networks. Physical Review Letters 99: 228701.
21. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. Science 311: 88–90.
22. Vázquez A, Oliveira JG, Dezsö Z, Goh KI, Kondor I, et al. (2006) Modeling bursts and heavy tails in human dynamics. Physical Review E 73: 036127.
23. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, et al. (2011) Small but slow world: How network topology and burstiness slow down spreading. Physical Review E 83: 025102.
24. Malmgren RD, Stouér DB, Motter AE, Amaral LA (2008) A poissonian explanation for heavy tails in e-mail communication. Proceedings of the National Academy of Sciences of the United States of America 105: 18153–18158.
25. Malmgren RD, Stouér DB, Campanharo ASLO, Amaral LA (2009) On universality in human correspondence activity. Science 325: 1696–1700.
26. Jo HH, Karsai M, Kertész J, Kaski K (2011) Circadian pattern and burstiness in human communication activity. arXiv:1101.0377.
27. Oliveira J, Vazquez A (2009) Impact of interactions on human dynamics. Physica A: Statistical Mechanics and its Applications 388: 187–192.
28. Min B, Goh KI, Kim IM (2009) Waiting time dynamics of priority-queue networks. Physical Review E 79: 056110.
29. Cho WK, Min B, Goh KI, Kim IM (2010) Generalized priority-queue network dynamics: Impact of team and hierarchy. Physical Review E 81: 066109.
30. Min B, Goh KI, Vazquez A (2011) Spreading dynamics following bursty human activity patterns. Physical Review E 83: 036102.
31. Stehlé J, Barrat A, Bianconi G (2010) Dynamical and bursty interactions in social networks. Physical Review E 81: 035101.
32. Zhao K, Stehlé J, Bianconi G, Barrat A (2011) Social network dynamics of face-to-face interactions. Physical Review E 83: 056109.
33. Holme P, Newman MEJ (2006) Nonequilibrium phase transition in the coevolution of networks and opinions. Physical Review E 74: 056108.
34. Iniguez G, Kertész J, Kaski KK, Barrio RA (2009) Opinion and community formation in coevolving networks. Physical Review E 80: 066119.
35. Kovanen L (2009) Structure and dynamics of a large-scale complex social network. Master's thesis, Espoo, Finland, Aalto University.
36. Lee SH, Kim PJ, Jeong H (2006) Statistical properties of sampled networks. Physical Review E 73: 016102.