The emergence and role of strong ties in time-varying communication networks

Márton Karsai\textsuperscript{1,2}, Nicola Perra\textsuperscript{1}, and Alessandro Vespignani \textsuperscript{*1,3,4}

\textsuperscript{1}Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston MA 02115 USA
\textsuperscript{2}BECS, School of Science and Technology, Aalto University, P.O. Box 12200, FI-00076
\textsuperscript{3}Institute for Scientific Interchange Foundation, Turin 10133, Italy
\textsuperscript{4}Institute for Quantitative Social Sciences at Harvard University, Cambridge, MA, 02138

March 26, 2013

Abstract

In most social, information, and collaboration systems the complex activity of agents generates rapidly evolving time-varying networks. Temporal changes in the network structure and the dynamical processes occurring on its fabric are usually coupled in ways that still challenge our mathematical or computational modelling. Here we analyse a mobile call dataset describing the activity of millions of individuals and investigate the temporal evolution of their egocentric networks. We empirically observe a simple statistical law characterizing the memory of agents that quantitatively signals how much interactions are more likely to happen again on already established connections. We encode the observed dynamics in a reinforcement process defining a generative computational network model with time-varying connectivity patterns. This activity-driven network model spontaneously generates the basic dynamic process for the differentiation between strong and weak ties. The model is used to study the effect of time-varying heterogeneous interactions on the spreading of information on social networks. We observe that the presence of strong ties may severely inhibit the large scale spreading of information by confining the process among agents with recurrent communication patterns. Our results provide the counterintuitive evidence that strong ties may have a negative role in the spreading of information across networks.

keywords: Network models, Data science, Information diffusion

1 Introduction

In the last ten years the access to high resolution datasets from mobile devices, communication, and pervasive technologies has propelled a wealth of developments in the analysis of large-scale networks\textsuperscript{1,2,3,4}. Particular efforts have been devoted to characterize how their structure influences the critical behaviour of dynamical processes evolving on top of them. This issue is extremely important to understand and model the spreading of ideas, diseases, informations, and many others dynamical phenomena\textsuperscript{5,6,7,8}. However, the large majority of the approaches put forth to tackle this subject

\textsuperscript{*}a.vespignani@neu.edu
uses a time-aggregated representation of network's interactions and neglect their time-varying nature. While this time scale separation is extremely convenient for practical reasons, it might introduce strong biases in the description of the processes. Indeed, the concurrency, and time ordering of interactions, even if the social network contains stable relationships, are crucial and may have considerable effects [9 10 11 12 13 14].

The characterization and modelling of time-varying networks are still open and active areas of research [15 16]. In this context, relational event-based network analysis enable to model network dependent, time-stamped event data [3] as well as human, and organizational interactions [17 18]. Appropriate dyadic level statistics govern the rate at which actors send out communications to their neighbors encoding traditional network structures as well as actor level attributes or even the history of actor level events for the sender. A simplification of this framework has been recently proposed by the activity-driven generative algorithm for time-varying networks [14]. This approach defines the activity potential, a time invariant function characterizing the agents’ interactions, and constructs an activity-driven model capable of encoding the instantaneous temporal description of the network dynamics. Within this simplified Markovian framework highly dynamical networks can be described analytically. This allows for a quantitative discussion of the biases induced by time-aggregated representations in the analysis of dynamical processes occurring on top of them [14 19 20]. However, this framework is memoryless, and it misses important features of real world systems. Indeed, each individual’s social circle has heterogeneous, and time-varying interactions. We can define them as strong links when they are repeated frequently as opposed to weak links signaling occasional interactions. The role of such diverse social ties in diffusion processes is extremely relevant [21], however a full understanding of their effects, explicitly considering the time-varying nature of interactions, is still missing.

In this paper we propose a generative model explaining the emergence of heterogenous ties within one’s social circle. We perform a thorough analysis of a large-scale mobile phone-call (MPC) dataset, which contains the records of time-stamped communication events of more than six millions individuals (for detailed description see Appendix A). We show that in this system the dynamics can be explained by introducing simple memory effects mathematically encoded by non-Markovian reinforcement processes. The introduction of this mechanism in the activity-driven model allows capturing the evolution of the egocentric network of each actor in the system, recovering also its global dynamics.

Along with the formulation of the generative model, we study the effect of intensity and time-varying interactions on a family of dynamical processes, namely rumour spreading models [22 23]. We assume that the time scales of the contact patterns evolution, and the spreading process are comparable. Interestingly, our findings clearly show that memory in the interaction dynamics hamper the rumour spreading process. Strong ties have an important role in the early cessation of the rumour spreading by favouring interactions among agents already aware of the rumour. The celebrated Granovetter conjecture that spreading is mostly supported by weak ties [40], goes along with a negative effect of the strong ties. In other words, while favouring locally the spreading, strong ties have an active role in confining the process for a time sufficient to its cessation. We validate the basic assumption, and modelling framework against results of data-driven simulations performed in the actual MPC time-varying network.

2 Results and Discussion

Here we focus on a prototypical large scale communication network where mobile phone users are the node of the network and the calls among them define the connectivity patterns. The common analysis
framework for such systems neglects the temporal nature of the connections in favour of time aggregated definitions of networks’ properties. In these descriptions, the degree \( k \) of a node represents the total number contacted individuals, while each link’s weight \( w \), the strength of the tie, indicates the total number of calls between the pair of connected nodes. The distributions of these quantities are shown in Fig.1a, and b. Interestingly, they are characterized by heavy-tailed behaviors. Although the time-aggregated network provides important information about the network structure, it does not provide information on the processes driving its dynamic. This issue is clearly exemplified in Fig.2a and b where two snapshots of the network at different times (covering few hours of activities in a town) capture temporal interactions of people that are instead not visible from the aggregated network representation (Fig.2c). The resulting structure, and dynamics of agents’s immediate social network is strongly heterogeneous as the social efforts of the ego are not distributed evenly towards his/her neighbors.

Figure 1: Distributions of the characteristic measures of the aggregated MPC network, and the activity-driven network. In panels (a), and (d) we plot the degree distributions. In panels (b), and (e) we plot the edge-weight distributions. Finally, in panels (c), and (f) we plot the node-activity distributions. In each figure grey symbols are assigning the original distributions while colored symbols are denoting the same distributions after logarithmic binning. In panels (d), (e), and (f) solid lines are assigned to the distribution induced by the reinforced process, while dashed lines denote results of the original memoryless process. Model calculations were performed with parameters \( N = 10^6 \), \( \epsilon = 10^{-4} \) and \( T = 10^4 \).

Egocentric networks were thoroughly investigated earlier in psychology and sociology [24, 25, 26], while some other characteristics have been recently mapped out with the availability of large-scale data [27, 28, 29, 30, 31]. Here we focus on the fine grained temporal behaviour of agents in order to identify some of the mechanisms which are driving the evolution, and dynamics of their egocentric networks (egonets). We tackle this problem by focusing on an important quantity, the activity rate \( a \), that allows describing the network evolution beyond simple static measures. It is defined as the probability of a node to be involved in an interaction at each unit time. The activity distribution is also heavy-tailed (see Fig.1c), but contrary to other measures such as degree and weight, it is a time invariant property of individuals [14]. It does not change by using different time aggregation time scales, and allows to encode crucial features of the instantaneous temporal description of networks’ dynamics into a generative modelling framework [14]. Here we extend this modelling framework by identifying, and modelling another crucial component: the role of memory in links’ evolution. We encode memory in a simple mechanism reproducing with great accuracy empirical data.

3
2.1 Egocentric networks dynamic

In order to identify the basic processes able to explain the structure of the instantaneous as well as the final integrated network, we analyze the growth of the egonet of each agent as a function of time. In particular, in one’s immediate social network we can distinguish between two types of social links. The first class describes strong, well established ties corresponding to repetitive, frequent interactions. The second class characterizes weak ties that are activated only occasionally. By looking at the network dynamics it is possible to assume that the interactions defining strong ties would appear earlier, and after most of those ties have been explored, only new weak ties are incrementally added to the egonet of each agent. Empirical evidence pointing out to this dynamics was recently observed in Ref. [32]. This intuition implies a particular dynamics in which the number of new connections of each node increases over time, but after all important links have been activated the network evolution slows down as new connections appear only sporadically. Note that here we are not entering a detailed discussion of additional dynamic processes such as changes in social status, aging, or permanent breaking of social ties that generally acts on different time scales. For the sake of a quantitative characterization, we measure the function \( p(n) \), which is the probability that the next communication event of an agent with \( n \) social ties will occur \textit{via the establishment of a new} \((n+1)\text{ social tie}\). To calculate these probabilities correctly we average them for users whose egonet has the same degree \( k \) at the end of the observation time, assuring averages are on the same number of individuals for each \( n \). We therefore measure the quantity \( p_k(n) \) for the egonets with the same degree \( k \) and \( n \leq k \). The empirical \( p_k(n) \) functions for different degree groups are shown in Fig.3 inset (colored symbols). The calculated probabilities are indeed decreasing with \( n \) for each degree class assigning a reducing velocity of the observed egocentric network evolution. The observed behaviour indicates that larger the egocentric network, the smaller the probability that the next communication event will be with someone who was not contacted before. The empirical observations confirm our hypothesis as it reflects the presence of memory; individuals remember their social ties and tend to repeat interactions on those already established ties.

The observed empirical behaviour of the egonet growth can be captured by a simple mechanism where we assume that the probability of having one interaction with someone who is already in the egocentric...
network is $n/(n + c)$. At the same time the probability having an interaction by establishing a new tie:

$$p(n) = 1 - \frac{n}{n + c} = \frac{c}{n + c}$$  \hspace{1cm} (1)$$

where $c$ is an offset constant depending on the degree class considered. One can fit the function of Eq.1 on the empirical data (solid lines in Fig.3 inset) and determine the corresponding constant $c$ for each degree group (see Table 1 in SM for the obtained values). Using the measured $c$ we can rescale the empirical $p_k(n)$ functions as

$$p_k(n/c) = \frac{1}{(n/c + 1)}$$  \hspace{1cm} (2)$$

and collapse the empirical data points of different degree groups on a single curve of the form of Eq.2 (as it is shown in Fig.3 main panel). This suggests that the same mechanism plays role during the egonet evolution of all individuals independently of their final number of connections. In the Supplementary Materials (SM) we provide analogous results considering the directed representation of the communication network.

2.2 Activity-driven network model with memory

We consider here the basic definition of the activity-driven network model \[14\]. In this approach $N$ disconnected nodes may entertain $m$ interactions according to a pre-assigned individual activity rate $a_i = \eta x_i$, where $x_i$ is drawn from a desired activity potential function distribution $F(x_i)$, and $\eta$ fixes the time scale of the process (see Appendix B.1). Motivated by empirical evidences \[14, 33, 34\] we assume heavy-tailed activity potential $F(x_i) \propto x_i^{-\nu}$ with exponent value $\nu = 2.8$, and without loss of generality we fix $\eta = 1$ (if not otherwise specified). In the basic activity-driven model the dynamics is memoryless (ML); i.e. at each time step all connections established previously are removed, and the new connections of each active node are established with randomly selected nodes with no memory of previous interactions.

Here we define a version of the model that considers a simple reinforcement process (RP) aimed at capturing a more realistic dynamical evolution for the local egocentric networks \[35, 36, 29\]. Namely, if a node with $n$ previously established social ties is activated at a given time step, with probability
Figure 4: Rumour spreading processes in (a) ML and (b) RP activity-driven networks. Node colors assign their states as ignorant (blue), spreader (red) and stifler (yellow) states. Node sizes, color, and width of edges represent the corresponding degrees and weights. The parameters of the simulations are the same for the two processes: $N = 300$, $T = 900$, $\lambda = 1.0$, and $\alpha = 0.6$. The process was initiated from a single seed with maximum strength.

$p(n) = c/(n + c)$ it will contact a node never contacted before by randomly selecting among the nodes of the system and establish a new social tie. Otherwise with probability $1 - p(n) = n/(n + c)$ it will interact with a node already contacted in previous time steps, thus reinforcing earlier established social ties. Memory is therefore introduced in the dynamics as each node keeps remembering the list of already established ties. Here we fix $c = 1$ for all the nodes and we leave the generalization of the model where this value is correlated with nodes properties for future studies (indeed we show in the SM how the emerging network properties are changing for different values of $c$).

A side by side comparison of a time-aggregated representation of the network structures generated by the ML and RP models (using the same parameters) are shown in Fig.4a and b. The ML dynamics (Fig.4a) induces a random aggregated network with a degree distribution $P(k) \propto k^{-\gamma}$ where $\gamma = \nu$ and edge weights are homogeneous [14]. This is also confirmed by larger scale simulation results depicted in Fig.4d and e (dashed lines). In case of the RP dynamics (Fig.4b) memory induces a considerably different structure by reinforcing recurrent communication on existing ties. This mechanism controls the emerging degree distribution as the evolution of egocentric networks become slower and simultaneously inducing heterogeneities in edge weights. These features are readily shown in Fig.4d, e, and f (solid lines) where a broad but skewer degree distribution reflects the reduced topological heterogeneities. The observed statistical properties as encoded by the measured distributions, matches the corresponding empirical measure in Fig.1a, b, and c. Numerical results suggest that the model degree distribution exponent obeys the $\gamma = 2\nu$ relationship with the corresponding activity exponent value if $c = 1$ (for further details see SM). Furthermore we find remarkable weight heterogeneities in the evolving RP network (see Fig.4e solid line) that captures extremely well real data contrary to the ML model (see Fig.4e dashed line). The RP dynamics not only induces realistic heterogeneities in the network structure, but also controls the evolution of the macroscopic network components. Indeed, due to the reinforced connection patterns the largest connected component (LCC) in RP networks evolves considerably slower compared to equivalent ML measures (for illustration see Fig.5a). This is a significant feature as no collective phenomena can evolve in a network on macroscopic scale faster then the LCC, thus affecting
the behaviour of dynamical processes in time-varying networks with memory.

2.3 Rumour spreading processes

One of the main advantages of the activity-driven class of models is that it is particularly convenient for the mechanistic simulation of dynamical processes co-evolving with the time-varying connectivity patterns of the network. Here we focus on the classic rumour spreading class of dynamical processes (for definition see Appendix B.2). In order to investigate the biases induced by simulating the rumour spreading on static time aggregated views of the network, we simulate this process in networks obtained integrating over time. We aggregate networks generated by ML and RP models (see Fig.5.b inset) and compare the results with their time-varying counterparts (see Fig.5.b main panel). The results obtained by the same parameters show striking differences between the velocity of spreading in the time aggregated views, and the activity-driven cases. The time for the rumour to reach a consistent fraction of nodes varies of orders of magnitudes in the two cases, with a very slow dynamics in the case of time-varying networks. These results evidence the significance of the temporal network approach as they change dramatically the critical regime of the investigated process. At the same time they highlights the possible limitations of studies performed on static structures.

![Figure 5](image-url)

Figure 5: In panel (a) we show the sizes of the largest connected components (LCC) as a function of time evolving in the aggregated ML and RP networks. Simulations are run with the same parameters for size $N = 10^5$. In panel (b) we show the stifler $r(t)$ density in rumour spreading simulations using ML (main panel, blue dashed line) and RP (main panel, purple solid line) processes with $N = 10^5$ and $\alpha = 0.6$ for $T = 10^5$. The rumour spreading processes were simulated with the same parameters on aggregated ML (inset, yellow dashed line) and RP (inset, brown solid line) networks integrated for $T$ times.

Another striking difference is observed in the final contagion densities in the ML, and RP simulations (Fig.5b main panel). In case of ML networks at the end of the rumour spreading $\sim 85\%$ of the network is aware of the rumour but in the RP case the final contagion proportion is only slightly more than $60\%$ of the total nodes. This hampering of the contagion process is also shown in Fig.4a and b where we simulated rumour spreading processes in ML and RP networks with the same parameters. The differences are straightforward not only in the evolving structure, but also in the level of contagion. In the RP case the rumour has spread only locally, while during the same time in the ML driven network the information reached a large fraction of nodes.

To investigate further the effects of the different activity-driven network models we performed large scale simulations using different initial conditions, and spreading rules. In each case we performed $10^3$ (or $10^4$ for smaller systems) simulations in identically parametrized ML and RP networks, where the process lasts at least $10^3$ steps. We evaluated each surviving process for $T = 5 \times 10^4$ time steps, and we measured the average final proportion of nodes aware of the rumour $\langle r_{eq} \rangle$. To highlight differences arising between the two dynamics we kept $\lambda = 1$ and calculated the $\langle r_{eq}^{RP}(\alpha) \rangle / \langle r_{eq}^{ML}(\alpha) \rangle$ ratios as the function of $\alpha$ infection rate. The network sizes considered are $N = 10^4$, and $10^5$. We initiated the spreading
from (i) the most active seed, (ii) one randomly selected seed or (iii) ten random seeds. Results in Fig.6 indicate marginal size effects but strong dependence on the initial conditions. All corresponding ratios are decreasing with $\alpha$, which highlights increasing differences between the fraction of population reached by the rumour in the two dynamics.

The largest differences are observed for single initial seed, specially in the case of the most active nodes (case i) red lines, and symbols in Fig.6. This numerical finding can be understood by considering that the rumour spreading process, and the reinforcement process are occurring on comparable time scales. As the applied reinforcement mechanisms induce recurrent interactions, they enhance the rumour spreading cessation by “pair annihilation” of nodes connected by strong ties. This effect is controlled by $\alpha$ and can induce up to $\sim 45\%$ relative difference in the population reached by the rumour in the case of the RP model.

To verify that the effect observed in the networks generated with the activity-driven network framework are present also in real communication networks we performed additional data-driven simulations. We have considered the actual MPC sequence of calls and simulate the rumour spreading by using this sequence as the time-varying network substrate (for more details see Appendix B.3). At the same time to directly contrast the role of memory and repeated interactions we defined a random null model. In this model we use the MPC sequence, keeping the caller of each event but select a callee randomly. This way we receive a sequence recovering the original activities, but with degrees and egocentric networks shuffled and inter-event correlations removed. The corresponding simulation results in Fig.7.a shows a striking difference in the speed of spreading and final density of stifler nodes. While for the shuffled case everyone becomes stifler at the end of the simulation, by using the original interaction sequences less than 40% of the network is aware of the rumour. This effect is even more evidenced in Fig.7.b where simulations were executed with the same parameters using the MPC and shuffled sequences. The ratio of corresponding stifler node densities were recorded after 182 days. Their relative difference is rapidly increasing and becomes several orders of magnitude large for larger $\alpha$ values. Different initial conditions are playing similar roles as in the activity-driven model processes. The effect of memory and repetitive interactions are the strongest if we initiate the rumour from the most active individual, while similar but weaker effects evolve if we select a single or multiple random seeds.

To show further evidences we measured the surviving probability $P_s(t)$ defined as the probability that a rumour spreading process survived (still contains nodes actively spreading the rumour) up to the time $t$ [38, 39]. Surviving probabilities of processes with different $\alpha$ are presented on Fig.7.c. The initial

![Figure 6: The $r_{eq}^{RP}/r_{eq}^{ML}$ ratios of average stifler densities in equilibrium. Simulations of models with sizes $10^5$ ($10^4$) are run with various initial conditions (see caption). Averages are calculated at $T = 5 \times 10^4$ for $10^3$ ($10^4$) realizations reaching equilibrium later than $T = 10^3$.](image-url)
Figure 7: In panel (a) we show the stifler $r(t)$ density in data-driven rumour spreading simulations using the MPC (purple solid line) and shuffled MPC (blue dashed line) sequences with $\alpha = 0.1$. Panel (b) depicts the $r_{MCN}/r_{MCN\ rand}$ ratios of average stifler densities in equilibrium. Simulations of panels (a) and (b) were run with various initial conditions (see caption) and averaged over $10^3$ realizations. $P_s(t)$ Surviving probability of rumour spreading processes initiated from single random seeds in (c) data-driven simulations using real MPC sequences and (d) simulations based on shuffled MPC sequences. Probability values of panels (c) and (d) were averaged over $10^4$ realizations.

scaling of $P_s(t)$ shows that generally the rumour can spread only locally due to repeated interactions occurring on strong links between the seed and its neighborhood. However, if the rumour survives this initial period and an outbreak takes place, then it will likely reach the whole population. A very different behaviour emerges if we remove the effect of memory and repeated interactions. The emerging differences are evident (see Fig.7(c)) as the initial effect of repeated interactions vanishes and all realizations survive until the rumour covers the whole network. Note that similar results were obtained for activity-driven model processes presented in the SM. This highlights the significant role of recurrent interactions via strong links as counterintuitively they play a bottleneck for the information propagation and they control the global outbreak of rumour spreading phenomena.

### 3 Conclusions

By analyzing a large scale longitudinal dataset of social interactions via mobile phone calls, we are able to provide a simple empirical characterization of the effect of memory in the dynamical evolution of the egocentric networks of individuals. Considering the empirical evidences we defined a simple reinforcement process that we used to integrate into a generative algorithm for time-varying networks with memory. The proposed network model mirrors many of the properties observed in the dataset and shows the spontaneous emergence of non-trivial connectivity patterns characterized by strong and weak ties. The analysis of rumour spreading processes on the synthetic time-varying networks clearly shows the biases introduced by using time-aggregated network representation in the study of dynamical processes. In particular, we see that neglecting the time-varying dynamics of the connectivity patterns changes both the time-scale, and final size of the rumour spreading process. In this context we also highlight the importance of strong ties responsible for constraining the rumour diffusion within localized groups of individuals. This evidence points out that strong ties may have an active role in weakening the spreading
of information by constraining the dynamical process in clumps of strongly connected social groups. The presented results underline the subtleties inherent to the analysis of dynamical processes in time-varying networks. No one-fits-all picture exist, and a classification of dynamical processes behaviour calls for a thorough analysis of each particular processes and networks considered. Furthermore, several extensions of the utilized framework of activity-driven networks are possible. Examples are node-node correlations, heterogeneous dynamics, and bursty behaviour of nodes. The present study thus offers potential avenues for the study of dynamical processes in time-varying networks in complex settings where the memory of agents plays a determinant role in the evolution of the connectivity patterns of the system.

This work has been partially funded by the NSF CCF-1101743 and NSF CMMI-1125095 awards. MK acknowledges support from EU’s 7th Framework Programs FET-Open ICTeCollective project (No. 238597). We thank A.-L. Barabási for the dataset used in this research.

A Dataset

The utilized dataset consists of 633,301,986,311 time stamped mobile-phone call (MPC) events recorded during 182 days with 1 second resolution between 6,243,322 individuals connected via 16,783,865 edges. The dataset was recorded by a single operator with 20% market share in an undisclosed European country (ethic statement was issued by the Northeastern University Institutional Review Board). To consider only true social interactions, and avoid commercial communications we used interactions between users who had at least one pair of mutual interactions.

B Model definitions

B.1 Reinforced activity-driven model

$N$ disconnected nodes are initially considered. Each one of them is assigned with an activity rate $a_i = \eta x_i$, where $x_i$ denotes the activity potential drawn from a desired $F(x_i)$ distribution ($x_i \in [\epsilon, 1]$), and $\eta$ is a rescaling factor to have $\eta \langle x \rangle N$ average number of active nodes per unit time. In order to explain the generative model process, let us consider a node $i$ that at time $t$ had $n$ different neighbors. At time $t+\Delta t$, the node $i$ is active with probability $a_i \Delta t$. If active, the node will connect $m$ other nodes selecting them with probability $p(n) = 1/(n+c)$ at random among the others $N-(n+1)$ never selected before, or with probability $1-p(n) = n/(n+c)$ among the $n$ previously selected. Even if the node is inactive it can receive contacts from other active nodes. At the next time step all the links in the networks are deleted, and the process is re-started. To simulate longer sequence of interactions we repeat the above steps in $T$ times. In this definition to track the actual neighbors, every node keeps remembering who they have connected earlier. This is different from the original model process where a node to connect is always selected randomly. For all of our calculations we fix the parameters $\eta = 1, \epsilon = 10^{-3}$ (if it is not noted otherwise), and $m = 1$ as calls take place between two people usually. We assume that the activity potentials is distributed as $F(x_i) \propto x_i^{-\nu}$ with exponent value $\nu = 2.8$ and the reinforcement function $p(n)$ takes $c = 1$ for each individuals.
B.2 Rumour spreading model

To simulate rumour spreading processes we use the classical model of Daley and Kendall [22], and define its co-evolution with the network dynamics as follows. Each node can be in three possible states; ignorant (I), spreader (S) or stifler (R). We denote their densities at each time $t$ as $i(t) = I(t)/N$, $s(t) = S(t)/N$, and $r(t) = R(t)/N$ accordingly. At $T = 0$ everyone is ignorant except the selected single or multiple seeds who are set to be spreaders. During the process we allow to spread any influence just via events of the activated users. At the time of an interaction the states of the connecting nodes can change by the following rules (independently of the direction of the actual link): (a) $I + S \overset{\lambda}{\rightarrow} 2S$ or (b) $S + R \overset{\alpha}{\rightarrow} 2R$ or (c) $S + S \overset{\alpha}{\rightarrow} 2R$. Here $\lambda$ and $\alpha$ are the transition rates into the states of spreader or stifler accordingly. In all measurement (if it is not noted otherwise) we set $\lambda = 1$ and use $\alpha$ as a parameter assuming that only their ratio matters for the spreading behaviour (supporting results are summarized in SM). Using these rules the system can reach an equilibrium state where all spreaders are vanished, $\partial_t i(t) = 0$, and $\partial_t r(t) = 0$.

B.3 Data-driven model

In data-driven simulations we initiated the rumour from a randomly selected call event of a randomly selected user in the MPC network runnig the process for the length of the recorded period. When a realization arrived to the last event of the sequence, we used a periodic temporal boundary condition as we continued the process with the first event of the sequence [37]. However, as the simulations were executed no longer than the recorded time period, no event was used twice during one simulation run.

References


The emergence and role of strong ties in time-varying communication networks

M. Karsai, N. Perra and A. Vespignani

4 Measures of egocentric network evolutions by directed communications

In the main text we disclosed strong memory effects in the interactions dynamics of an ego and his/her social circle. To complete our analysis here we repeat all the measurements separately for directed outgoing and incoming call sequences. The characteristic functions as $P(k)$ and $P(a)$ of directed communication are very similar to the undirected case as it is evidenced in Fig. 8a and b. These counts remain broadly distributed and scaling similarly in all three cases.

![Figure 8: Characteristic distributions of networks aggregating directed communications. (a) In- (red) and out (green) degree distributions are compared to the overall undirected $P(k)$ (grey). (b) The incoming and outgoing activity distributions are shown together with the undirected $P(a)$ (with colors similar to figure a).](image)

To investigate the presence of memory, we categorize every events of an individual into two groups. One group contains actions, which evolve on a link where other events have taken place earlier. Events in this group does not increase the degree of the ego but contribute to the edge weight and activity potential. Events belonging to the other group evolve between the ego and someone else who he/she has never connected before. These events induce links with unit weight, which increase the ego's degree and also incremental to his/her activity. By using this categorization we can measure a conditional probability $p(n)$ that the next event of an individual will be towards one of his $n$ already existing neighbours, or with a new person who he/she has never called before. We measure this probability for different degree groups. We select people with $k$ number of neighbours where $k_{\text{min}} \leq k < k_{\text{max}}$ and calculate the $p(n)$ function for $n \leq k_{\text{min}}$ to assure that each calculated probability value is extracted from the same number of users belonging to the actual group.

We perform this measurement for directed and undirected communication. Results of undirected communications are reported in the main text while for directed communication sequences are shown in Fig. 9a and b. They all present very similar behaviour and can be fitted with functions in the form of
Figure 9: Inset: $p_k(n)$ functions for (a) outgoing and (b) incoming call sequences calculated for nodes in different degree groups. We also show fits of $p_k(n)$ functions (defined in Eq.1 in the main text) using different $c$ constant values (see Table 1). Main panel: scaling of $p_k(n)$ and fitting with the universal function defined in Eq.2 (see main text).

Eq.1 (see in main text). Fitting results for outgoing and incoming calls are shown in Fig.9.a and b inset, while their re-scaling (as Eq.2 in main text) is depicted in Fig.9.a and b main panels accordingly. The corresponding fitted $c$ constants of all directed and undirected sequences are summarized in Table 1.

Table 1: The fitted $c$ constants and $s_e$ standard error values for the observed and analytical $p(n)$ functions for different degree groups in undirected, outgoing and incoming communication sequences.

<table>
<thead>
<tr>
<th>$k_{min}..k_{max}$</th>
<th>$c \pm s_e$ (out)</th>
<th>$c_{out} \pm s_e$</th>
<th>$c_{in} \pm s_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>0.444 ± 0.053</td>
<td>0.553 ± 0.05973</td>
<td>0.475 ± 0.0474</td>
</tr>
<tr>
<td>4.7</td>
<td>0.584 ± 0.065</td>
<td>0.713 ± 0.06143</td>
<td>0.648 ± 0.0492</td>
</tr>
<tr>
<td>8.16</td>
<td>0.814 ± 0.070</td>
<td>0.964 ± 0.06003</td>
<td>0.970 ± 0.04415</td>
</tr>
<tr>
<td>16..32</td>
<td>1.264 ± 0.056</td>
<td>1.571 ± 0.05505</td>
<td>1.551 ± 0.03262</td>
</tr>
<tr>
<td>32..64</td>
<td>1.928 ± 0.059</td>
<td>2.469 ± 0.05031</td>
<td>2.429 ± 0.03998</td>
</tr>
<tr>
<td>64..128</td>
<td>3.077 ± 0.109</td>
<td>4.238 ± 0.2622</td>
<td>4.017 ± 0.2525</td>
</tr>
<tr>
<td>128..256</td>
<td>10.913 ± 0.826</td>
<td>14.694 ± 1.735</td>
<td>15.836 ± 2.278</td>
</tr>
</tbody>
</table>

These results evidence that similar memory effects can be detected in directed and undirected communication sequences. They all indicate that the larger one’s observed personal social network the larger the probability that he/she will make (receive) a call towards (from) someone who is already in his/her egocentric network.

5 Degree evolution of reinforced activity driven networks

As it has been shown earlier [1] the degree distribution of the integrated structure of a memoryless (ML) activity-driven network follow the same functional form as the activity distribution. If the activity distribution scales as $P(a) \sim a^{-\nu}$ then the degree distribution should be $P(k) \sim k^{-\gamma}$ with $\nu = \gamma$. This was shown analytically in [1] and is confirmed by simulation results in Fig.10.a where $P(k)$ evolves with the same exponent value $\gamma = \nu = 2.8$ as $P(a)$ was characterized. The distributions there also show strong finite-size effects. If the simulated network size is finite and we increase the integration time the networks become more and more connected approaching to a fully connected graph. This effect causes that small degree nodes are not presented for large integration time.

Networks generated by reinforced activity-driven processes (RP) also evolve with heterogeneous degrees and with similar finite size effects as above. However, in this case the relation between $\gamma$ and $\nu$ is
somewhat different. Here the egocentric network evolution is controlled by the reinforced interactions. Egonets evolve slower as interactions of agents are reinforced to take place on already established links. This effect induce reduced degree heterogeneities with exponents $\gamma$ larger than the one characterizing the activity distribution. It is visible in Fig. 10 where $\gamma$ of the evolving ML and RP networks are strikingly different even the activity exponent $\nu = 2.8$ and any other parameters were chosen to be the same for both processes.

Figure 11: Relation between node strength and degree in ML and RP model networks. (a) Degree-strength correlations for ML model in networks generated by different $\nu$ activity exponents. The fitted function is linear in $k$. (b) Degree distribution of the same ML networks. (c) Degree-strength correlations for RP model in networks generated by different $\nu$ activity exponents. The fitted function is $k^2$. (d) Degree distribution of the same RP networks. Common parameters of the simulations: $N = 1,000,000$, $m = 1$, $\epsilon = 0.001$, $T = 10,000$. 

Figure 10: Degree distributions of (a) ML and (b) RP activity networks integrated through $T = 10^4$, $10^5$ and $10^6$ time steps. The fitting power-law exponent values are (a) $\gamma = 2.8$ and (b) $\gamma = 5.6$. Common parameters of the simulations were $N = 100,000$, $m = 1$, $\epsilon = 0.0001$. 
In the activity driven framework the degree of a node $i$ at time $t$ is decomposed into two parts as
\[ k_i(t) = k_i^{\text{out}}(t) + k_i^{\text{in}}(t) \] (3)
where $k_i^{\text{out}}$ is the number of other nodes whom the node $i$ connected, while $k_i^{\text{in}}$ is the number of other nodes who connected node $i$ up to time $t$. Similar to degrees, the probability to have a degree $k$ of a node $i$ at time $t$ can be decomposed in two terms as:
\[ P(t, k_i) = P_{\text{out}}(t, k_i) + P_{\text{in}}(t, k_i) \] (4)
where for an RP network (when $m = 1$) the two probabilities can be written as
\[ P_{\text{out}}(t, k_i) = a_i \left[ \frac{k_i - 1}{k_i} P(t - 1, k_i - 1) + \frac{k_i - 1}{k_i + 1} P(t - 1, k_i) \right] \] (5)
and
\[ P_{\text{in}}(t, k_i) = \sum_{i \neq j} a_j \frac{1}{k_j} \frac{1}{k_j + 1} \frac{1}{N - k_j + 1} P(t - 1, k_j) P(t - 1, e_{ij} \notin E_{t-1}) \] (6)
where $E_t$ denotes the actual edge list at time $t$. These equations could provide us the relation between $\gamma$ and $\nu$, however no closed analytical solution has been found so far.

Another way to estimate the relation between the activity and degree exponents in evolving RP networks is by directly measuring this correlation in large scale numerical simulations. In case of ML processes we have seen that $P(a) \sim P(k)$ [1] thus a $a \sim k$ linear correspondence should be apparent. By definition the $a$ activity rate of a node is proportional to its $s$ strength, the total number of events the node participated during the process, thus for the ML case $s \sim k$ relations should be also satisfied. This correlation is confirmed in Fig.11.a for several $\nu$ exponent values, where the correlation between $k$ and $s$ is apparent to be linear and in Fig.11.b where the degree distributions evolved with exponent satisfying the relation
\[ \gamma = \nu. \] (7)

By following the same train of thought, if we measure the same correlation for RP networks it should also disclose the dependency between the actual activity and degree exponents. Calculations in Fig.11.c indicates a more dispersed distribution of correlation values between $s$ and $k$, however it suggest that their relation can be characterized as $a \sim k^2$ independently from the $\gamma$ exponent value. This dependency provides a relation between the degree and activity exponents as
\[ \gamma \simeq 2\nu \] (8)
which is confirmed in Fig.11.d. The estimated $\gamma$ exponents fit well the corresponding degree distributions generated by RP dynamics using different $\nu$. In this way we could determine numerically the relation between the activity and degree distributions for reinforced processes in case of power-law distributed activities.

5.1 Degree and reinforcement dependence

In reinforced activity-driven networks the decision of an agent to establish a new or reinforce an already existing connection is driven by a $p(n)$ probability. Earlier we showed that this probability can be well
approximated with a simple analytical form as

$$p(n) = \frac{c}{n + c}$$

(9)

where only the $c$ parameter depends on the activity and degree of the actual agent. Even in our model calculations (for simplicity) we fixed $c = 1$ for every agent we remark that the evolving structural heterogeneities are depending on the choice of $c$. If $c \to 0$ the probability of calling a new friend goes to 0, which reduce the emerging degree differences but increase the evolving weight heterogeneities. In the limiting case $P(k)$ becomes exponentially distributed while $P(w) \sim P(a)$ as it is shown in Fig. 12a and b for $c = 10^{-5}$. On the other hand if $c \to \infty$ then $p(n) \to 1$ and the model approaches the memoryless activity-driven model process. In this limit $P(w)$ becomes exponential distribution and $P(k)$ takes the same functional form as $P(a)$ as it is depicted in Fig. 12a and b for $c = 10^5$. This way by varying $c$ one can control the emerging structural heterogeneities. One could even devise a model where $c$ follow a specific correlations with the activity of the actual agent, however, we let this kind of model extensions to be the subjects of future studies.

6 Spreading rate dependencies

During our simulations of rumour spreading processes we assumed that the actual values of $\lambda$ and $\alpha$ not, but their relative values matter for the equilibrium contagious level. To support this assumption we repeated some simulations of spreading processes with ordinary parameters and single random seeds. Here, instead of fixing $\lambda$ to unity, we performed measurements with $\lambda = 0.8, 0.6, 0.4$ and 0.2 values and record the $r_{RP}^\alpha(\alpha)/r_{ML}^\alpha(\alpha)$ curves in each case with $\alpha$ values ranging from 0 to $\lambda$.

The results depicted in Fig. 13 demonstrate that only the relative values of the two rates matter for the equilibrium contagious level. Here curves corresponding to $\lambda = 1.0, 0.8, 0.6$ and 0.4 are very similar with differences only due random fluctuations (all of them were averaged over 1000 realizations). The largest discrepancy appears with $\lambda = 0.2$ for small $\alpha$ values. This is because if $\lambda$ is small, the rumour spreads very slowly and even it would reach the same contagious level in equilibrium as any corresponding processes, it takes much longer time to reach this state. Since the time window was fixed to $T = 50,000$ for every simulations, some processes could not arrive to the final equilibrium state which induced the discrepant scaling of $r_{RP}^\alpha(\alpha)/r_{ML}^\alpha(\alpha)$ ratios for smaller $\alpha/\lambda$ values.
Figure 13: The $r_{RP}(\alpha)/r_{ML}(\alpha)$ curves for rumour spreading processes with various $\lambda$ infection rate values. Simulations were performed for ML and RP processes with $\lambda = 1.0, 0.8, 0.6, 0.4$ and 0.2 with relative $\alpha/\lambda$ values ranging between 0...1. Results were averaged over 1000 surviving simulations (see main text) with parameters $N = 10,000$, $m = 1$, $\epsilon = 0.001$, $T = 50,000$.

7 Surviving probability

The surviving probability is defined as the probability that a system still contains agents in the spreader state at time $t$ (in other words the rumour survived up to $t$). This probability was measured in the MPC data-driven simulations and results were reported in the main text. Here we repeated these measurements for rumours spreading in activity-driven networks.

Figure 14: $P_s(t)$ Surviving probability of rumour spreading processes driven by (a) reinforced and (d) memoryless activity-driven processes. Probability values were calculated on networks with parameters $N = 10^5$, $m = 1$, $\epsilon = 0.001$, $T = 50,000$ for rumours initiated from a single random seeds. Results are averaged over $10^2$ realizations.

If the network evolution is driven by reinforced interaction processes (see Fig.14a) we recover the same effect what was observed during the data-driven simulations (see Fig.7.c in the main text). Repetitive interactions and memory play apparent roles in the early stage of the spreading processes as the rumour may die out shortly after its initiation and could spread only locally. This behaviour can be concluded from Fig.14a (and simultaneously from Fig.7.c in the main text) where the surviving probability rapidly decrease in the initial time regime for larger $\alpha$ values. After the rumour survives the initial stage, it spreads globally and reach a considerable fraction of the network.

On the other hand if the activity-driven process is memoryless (or equivalently if the data-driven
spreading is evolving on shuffled event sequences) no repetitive interactions effect the initial temporal regime of the process and the rumour spreads always globally as it is evidenced Fig.14b (and in Fig.7.d in the main text where). This qualitative match between the model process and the data-driven simulations provides further prove that our model captures the role of memory and reinforcement processes in a consistent way.

References