Small but slow world: How network topology and burstiness slow down spreading

M. Karsai,[1,*] M. Kivelä,[1] R. K. Pan,[1] K. Kaski,[1] J. Kertész,[1,2] A.-L. Barabási,[2,3] and J. Saramäki[1]

[1]*BECS, School of Science and Technology, Aalto University, P.O. Box 12200, FI-00076, Finland*
[2]*Institute of Physics and BME-HAS Condensed Matter Group, BME, Budapest, Budafoki út 8., H-1111, Hungary*
[3]*Center for Complex Networks Research, Northeastern University, Boston, Massachusetts 02115, USA*

While communication networks show the small-world property of short paths, the spreading dynamics in them turns out slow. Here, the time evolution of information propagation is followed through communication networks by using empirical data on contact sequences and the susceptible-infected model. Introducing null models where event sequences are appropriately shuffled, we are able to distinguish between the contributions of different impeding effects. The slowing down of spreading is found to be caused mainly by weight-topology correlations and the bursty activity patterns of individuals.

Most complex physical, biological, and social networks show *small-world* property, where the average shortest path length is strikingly short compared to the network size [1]. This means that there is at least one short path between any two nodes that should give rise to rapid transmission of influence. However, dynamic phenomena on networks [2], such as spreading of pandemics, electronic viruses, and information, follow their own pathways, which are not necessarily topologically efficient [3,4]. Spreading on real small-world networks turns out to be surprisingly slow, e.g., new infections are reported years after the emergence of a new computer virus or the introduction of an antivirus [5]. Here we aim at resolving this puzzle. For issues such as strategies and timings of vaccination, improvement of information diffusion, and the slow decay of activity of computer viruses, it is crucial to understand the role of the underlying network and temporal activity patterns in the spreading dynamics.

The dynamics of spreading is commonly studied with SI, SIR, or SIS models [6] on static networks or in mean field, where the letters in the acronyms refer to the different states (Susceptible, Infectious, or Recovered) of individuals and the dynamics is defined by changes of these states due to the influence of others. These models lead to rapid, exponential growth of the epidemic at early stages of spreading, while the dynamics at later stages depend on the model and network. For the SI process, the infection spreads until the whole system reachable from initial conditions is infected, with exponential slowing down toward the end. For the SIR process, competing effects set in and the spreading may remain local or percolate through the system while the SIS process has more complex dynamics.

While these results capture some of the qualitative features of real-world processes, the heterogeneity of the systems limits their applicability. First, the interactions of real-world systems span networks by broad distributions of node connections and mesoscopic features in the form of communities with dense internal and sparse external connectivity. Second, interaction intensities vary and are closely coupled to network topology.

Third, the daily cycle and bursty character of interaction events give rise to important temporal inhomogeneities.

Some aspects of these features have already been studied. For static networks, it is known that spatial structure has an effect on epidemics (see, e.g., [7,8]), and community structure slows down information diffusion due to trapping in dense regions [9–11]. There is an intimate relation between inhomogeneous link weights and network topology in social and communication networks [12,13]: Links within communities are strong, while links between them are weak. This Granovetter-type structure enhances the trapping effect of the communities, leading to additional slowing down of spreading [13].

The bursty nature of human interactions has received particular interest and it has turned out that the corresponding activity patterns are usually non-Poissonian, often power-law correlated (see [14]). The effect of bursty dynamics on spreading has been approached using empirical data together with approximate analytical models [15,16]. In Ref. [15], computer worm spreading was studied using email logs and the SI model, and it was found that the non-Poissonian interevent time distribution leads to slow spreading in the late stages of the process. Slow spreading was also observed in Ref. [16], where an internet viral marketing experiment was carried out and modeled as a branching process in the nonpercolating regime. It was also argued that on the contrary, in the percolating regime, broad interevent time distributions should give rise to faster spreading.

In this Rapid Communication, we study the problem of spreading dynamics in its full complexity, using time-stamped event data on human communication networks and the SI model. We apply null models on these event sequences and show that spreading slows down due to simultaneous effects of structural and temporal correlations.

For the event sequences, we have used the following data: a) Mobile phone data from a European operator (national market share $\sim 20\%$) with $\sim 3.25 \times 10^8$ time-stamped voice call records over a period of 120 d. We have only retained links with bidirectional calls within the largest connected component (LCC) of the aggregated call network (MCN), yielding $N = 4\,572\,735$ nodes, $L = 9\,055\,944$ links, and $306\,218\,217$ calls. We define link weights as the number of

_____
*marton.karsai@aalto.fi

calls between two users. The network is sparse (average degree $\langle k \rangle = 3.96$) showing small-world property with an average shortest path length of $\langle l \rangle = 12.31$. b) Mobile call data from the Reality Mining project [19] (RM), where the LCC consists of 59 users and 93 edges with 2293 calls over ~9 months. c) Email logs [20] forming a network with the LCC having 2993 nodes and 28 843 edges for 202 687 events over 83 days. Here communications are directed and thus the nodes belong to the strongly connected component (SCC) where all nodes are reachable from each other, or the IN- or the OUT-component.

We study the SI spreading dynamics with simulations using the event sequences so that an infected individual infects a susceptible one at time t, if there is an event between them. For the events, we use records of the times and participants of calls, and the times and addresses of emails. Calls are one-to-one communication and enable *bidirectional* exchange of information, while emails may have multiple addresses and the information flow is *directed*. Hence for calls, if either participant is infected he/she infects the susceptible one, whereas for emails, transmission is from the sender to the recipient(s). We initiate simulations by infecting a randomly chosen node at a randomly chosen event with the spreading quantity (information, rumor, or virus) and set all other nodes susceptible. Then the spreading dynamics is simulated by using temporally periodic boundary conditions (i.e., repeating the event sequence) until the set of reachable nodes is exhausted. We record the prevalence, i.e., the fraction of infected nodes $\langle I(t) \rangle / N$ as a function of time averaging over 10^3 initial conditions and the time to full prevalence t_f. For the email network, we start the spreading process from a node in IN or SCC and iterate the process until all nodes in SCC and OUT are infected.

To gain insight into the effects of different correlations, we employ null models where the original event sequences are randomized. These are defined so that in each null model, some of the correlations are separately destroyed: community structure (C), weight-topology correlations (W), bursty event dynamics on single links (B), and event-event correlations between links (E). In addition, the overall event frequencies follow a daily pattern (D), with decreased night-time activity and some day-time peaks (see inset in Fig. 3) The null models are as follows, with the letters indicating retained correlations (Table I):

(i) DCWB (*equal-weight link-sequence shuffled*): Whole single-link event sequences are randomly exchanged between links having the same number of events. Temporal correlations

TABLE I. Correlations retained in different null models. D: daily pattern, C: community structure, W: weight-topology correlations, B: bursty single-edge dynamics, E: event-event correlations between edges. Rightmost column: Mean times to reach 20% prevalence and the errors of the mean values.

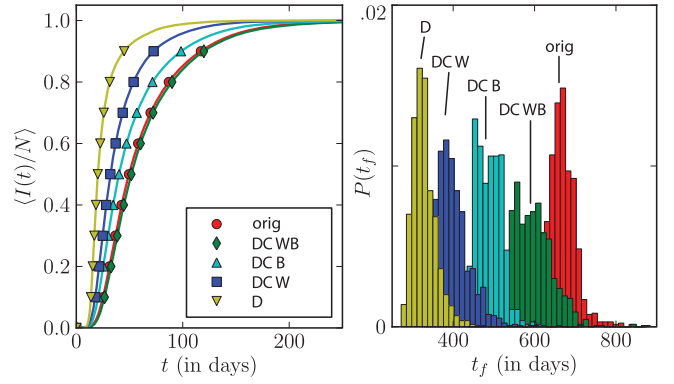| Event Sequence	D	C	W	B	E	$\langle t_{20\%} \rangle$ (days)
Original	√	√	√	√	√	38.24 ± 0.74
Equal-weight link-seq. shuf.	√	√	√	√		41.20 ± 0.83
Link-sequence shuffled	√	√		√		30.67 ± 0.55
Time shuffled	√	√	√			26.48 ± 0.52
Configuration model	√					17.23 ± 0.22

FIG. 1. (Color online) (Left) Average fraction of infected nodes $\langle I(t)/N \rangle$ at each points in time for the original event sequence (○) and null models: equal-weight link-sequence shuffled DCWB (△), link-sequence shuffled DCB (△), time-shuffled DCW (□), and configuration model D (▽). (Right) Distribution of full prevalence times $P(t_f)$ due to randomness in the initial conditions.

between links are destroyed. (For large weights we did binning with 2–3 weight values.)

(ii) DCB (*link-sequence shuffled*): Whole single-link event sequences are randomly exchanged between randomly chosen links. Event-event and weight-topology correlations are destroyed.

(iii) DCW (*time-shuffled*): Time stamps of the whole original event sequence are randomly reshuffled. Temporal correlations are destroyed.

(iv) D (*configuration model*): The original aggregated network is rewired according to the configuration model, where the degree distribution of the nodes and connectedness are maintained but the topology is uncorrelated. Then, original single-link event sequences are randomly placed on the links, and time shuffling as above is performed. All correlations except seasonalities like the daily cycle are destroyed.

Figure 1 displays the results for the MCN. In all cases the spreading is slow, with the full prevalence times t_f being several hundred days. It is clear that both topological and temporal correlations slow down the spreading. It is the fastest when all correlations except the daily patterns are destroyed (configuration model, D). Switching on the community structure and associated weight-topology correlations (DCW) slows down the spreading strongly, as expected due to bottlenecks caused by weak links between communities and the broad distribution of link weights [13,18]. However, comparing this with the DCB null model indicates that bursty single-edge dynamics (B) has an even stronger slowing-down effect than weight-topology correlations (W). Finally, including all except event-event correlations (DCWB) gives rise to spreading dynamics very close to the original event sequence (DCWBE). Here for early times, the DCWB spreading is slightly slower than the original one (see Table I) since temporal correlations (E) between adjacent edges have minor accelerating effect due to easy reachability within the community where the spreading begins, while for long times bottlenecks appear and event-event correlations slow the process down.

The slowing down effects can be characterized by the average times $\langle t_{20\%} \rangle$ to reach 20% prevalence, see Table I. The difference between the original and the fastest model is
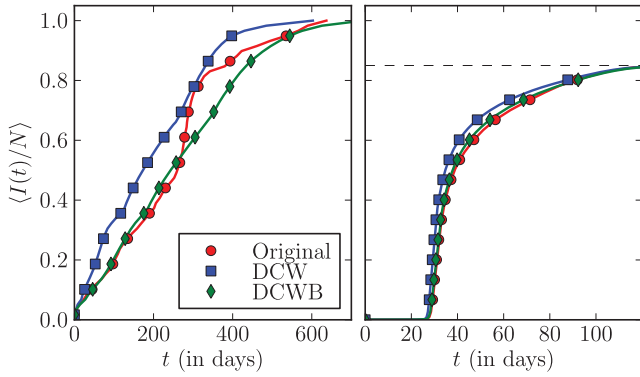
FIG. 2. (Color online) Spreading dynamics in the Reality Mining (left) and email networks (right), for the original event sequence (○) and null models: DCW (□) and DCWB (◇). In the email network, the spreading process is directed. The maximum prevalence is limited to the total fraction of the SCC and the OUT component (∼85%).

∼21 days, i.e., a factor ⩾2. Similarly for the 100% prevalence this factor also ⩾2 (∼342 d), showing that the effects of correlations are consistent for the duration of the whole process and for individual runs. As for the effect of the random initial conditions, the small error of mean values in Table I show that the mean curves (Fig. 1) characterize the overall behavior well. The effect of initial conditions are demonstrated in Fig. 1, where the distributions are clearly separable at full prevalence.

Results for the Reality Mining mobile call network and for the email logs are shown in Fig. 2, with the DCW and DCWB null models; the outcome is qualitatively similar with that of MCN. However, there are certain differences. In the small and sparse RM network, successive calls to many people within a short time period by a hub give rise to a steep prevalence rise. Such behavior is a one-off event and the effect is destroyed in the null models. In the email network, very high-degree hubs sending frequent emails give rise to rapid spreading once they are reached. This effect is conserved in the null models.

The daily activity pattern, i.e., variation in overall communication frequency by the hour, is retained in every null model that is based on randomizing the original event sequence. In [21], it was suggested that natural periodicities, such as the daily cycle, are responsible for the fat-tailed waiting time distributions. In order to evaluate the impact of the daily pattern on the spreading speed, we carried out simulations where the aggregated MCN was used as the network. Events were generated on its links by two Poisson processes that conserve link weights: a homogeneous Poisson process, and a process whose instantaneous rate follows the daily pattern as calculated from the call statistics on hourly basis (see inset in Fig. 3). The SI dynamics for both cases are shown in Fig. 3. The difference between the two curves is negligible, demonstrating that the daily pattern has only a minor impact on the spreading speed. This, together with the observation that temporal correlations do have a significant decelerating effect on spreading strongly indicates that there are important, non-Poissonian correlations in the system beside the daily type cycles.

The non-Poissonian, bursty character of event sequences is clearly demonstrated by the fat-tailed distribution of single-link interevent times for the MCN, see Fig. 4. In order to
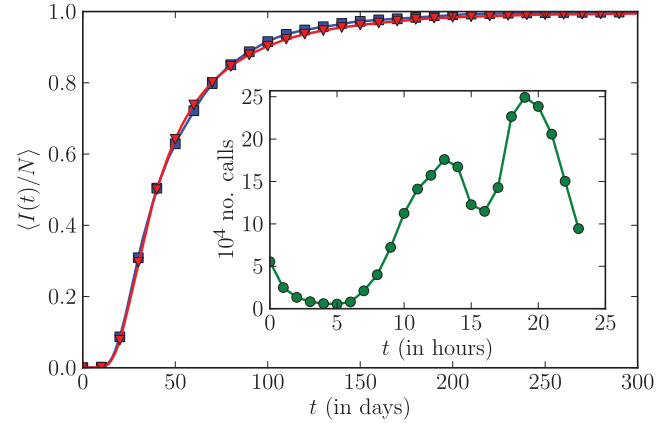

FIG. 3. (Color online) Spreading dynamics as obtained from a Poissonian event-generating model on the aggregated MCN, with daily pattern (□) and without (▽). Link weights were taken into account and the curve with the daily pattern is comparable with the DCW null model. Inset: the average daily pattern as observed for the MCN event sequence with binning by the hour. The continuous line is to guide the eye.

exclude the possibility that the fat tail in the interevent time distribution is only due to the broad weight distribution as suggested in [21], we calculated the distributions for binned weights and obtained satisfactory scaling with the average interevent time, same as [17]. We find that the distribution can be fitted by a power law with exponent 0.7 over 3.5 decades, followed by a fast decay. The scaling breaks down for small interevent times, where a peak in the distribution at ∼20 s is found, which is due to event correlations between links. The power law indicates non-Poissonian bursty character of the events. Both the characteristics vanish for the time-shuffled null model and the interevent time is well described by an
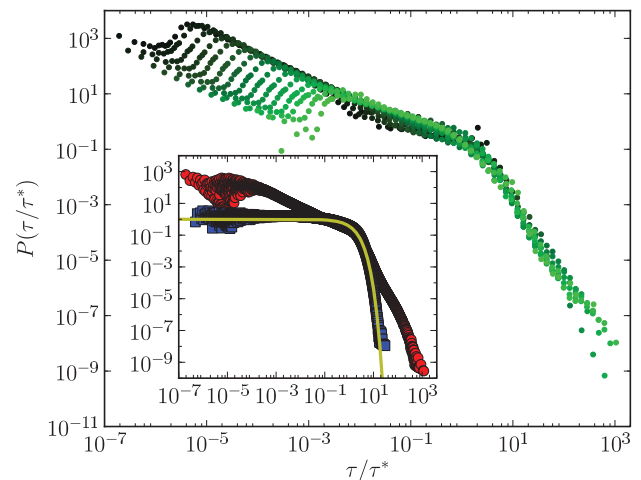

FIG. 4. (Color online) Scaled interevent time distributions for the MCN. Edges were log-binned by weight and for every second bin the interevent time distribution of the events occurring in the corresponding bin is shown, scaled by the average interevent time of that bin τ^* (larger τ^* darker the color). Inset: scaled inter-event time distributions for the original (○) and for the time-shuffled events (□). An exponential density distribution with average value of 1 is shown as a light (yellow) line.

exponential function (see inset of Fig. 4), i.e., the process is Poissonian.

The effect of burstiness on the spreading speed can be easily demonstrated with the following single-link calculation. Let us denote the average time for the infection to spread through a link (the residual waiting time) by $\langle \tau_R \rangle$, and assume that one of the nodes gets infected at a uniformly chosen random time. Similarly as in [15,16] we calculate $\langle \tau_R \rangle$ for a given interevent time distribution $P(\tau)$. For simplicity, we consider how the burstiness introduced by a continuous power-law distribution of interevent times $P(\tau) \sim \tau^{-\alpha}$ affects the average infection times when compared to a Poisson process. If we fix the average interevent time (and thus the number of events for a long observation period), the ratio of average infection times is $r = \langle \tau_{R,\text{power-law}} \rangle / \langle \tau_{R,\text{poisson}} \rangle = \frac{(\alpha-2)^2}{2(\alpha-1)(\alpha-3)}$ for $\alpha > 3$. Now r is decreasing with α, $r < 1$ when $\alpha > 2 + \sqrt{2} \approx 3.4$, and r goes to infinity at $\alpha = 3$. This indicates that the burstiness characterized by power law distributions with slow decay has a decelerating effect on spreading with respect to the Poisson process with the same mean. However, if the decay is fast enough, i.e., the second moment of the power-law distribution is smaller than that of the Poisson distribution, we see acceleration. This mean field type of reasoning has its limitations.

Nevertheless it illustrates the mechanisms of slowing down because of bursts: the residual waiting time increases because the chance for long waiting times after getting infected increases.

In conclusion, we have studied the effects of different topological and temporal correlations on spreading in complex communication networks. Using time-stamped event data and appropriately prepared null models we have managed to quantitatively distinguish between different contributions to the slowing down of spreading. We have shown that the main contributions are (i) the community structure and its correlation with link weights and (ii) the inhomogeneous and bursty activity patterns on the links. Somewhat surprisingly, the daily pattern and event correlations between links seem to play only a minor role in the overall spreading speed. Finally, we believe that our null models can be generally applied to investigate the effects of temporal and structural correlations on dynamic processes on networks.

[1] M. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2006); M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[2] A. Barrat, M. Barthelémy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).

[3] P. Holme, Phys. Rev. E **71**, 046119 (2005).

[4] D. Brockmann, L. Hufnagel, and T. Geisel, Nature **439**, 462 (2006).

[5] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Oxford University Press, Oxford, 2004).

[6] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Science Publications, Oxford, 1992).

[7] M. J. Keeling, Proc. R. Soc. London B **266**, 859 (1999).

[8] K. T. D. Eames, Theor Popul Biol **73**, 104 (2008).

[9] R. Lambiotte, J.-C. Delvenne, and M. Barahona, e-print arXiv:0812.1770 (2008).

[10] R. Toivonen *et al.*, Phys. Rev. E **79**, 016109 (2009).

[11] P. J. Mucha *et al.*, Science **328**, 876 (2010).

[12] M. Granovetter, Am. J. Sociol. **78**, 1360 (1973).

[13] J.-P. Onnela *et al.*, Proc. Natl. Acad. Sci. USA **104**, 7332 (2007).

[14] A.-L. Barabási, *Bursts: The Hidden Pattern Behind Everything We Do* (Dutton Books, New York, 2010).

[15] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási, Phys. Rev. Lett. **98**, 158702 (2007).

[16] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).

[17] J. Candia *et al.*, J. Phys. A: Math. Theor. **41**, 224015 (2008).

[18] J.-P. Onnela *et al.*, New J. Phys. **9**, 179 (2007).

[19] N. Eagle, A. Pentland, and D. Lazer, Proc. Natl. Acad. Sci. USA **106**, 15274 (2009).

[20] J. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. USA **101**, 14333 (2004).

[21] R. D. Malmgren *et al.*, Proc. Natl. Acad. Sci. USA **105**, 18153 (2008); **325**, 1696 (2009).