Epidemic Spreading in Non-Markovian Time-Varying Networks

Kaiyuan Sun,1 Andrea Baronchelli,2 and Nicola Perra1

1Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston MA 02115 USA
2Department of Mathematics, City University London, Northampton Square, London EC1V 0HB, UK

(Dated: April 4, 2014)

Most real networks are characterized by connectivity patterns that evolve in time following complex, non-Markovian, dynamics. Here we investigate the impact of this ubiquitous feature by studying the Susceptible-Infected-Recovered (SIR) and Susceptible-Infected-Susceptible (SIS) epidemic models on activity driven networks with and without memory (i.e., Markovian and non-Markovian). We show that while memory inhibits the spreading process in SIR models, where the epidemic threshold is shifted to larger values, it plays the opposite effect in the case of the SIS, where the threshold is lowered. The heterogeneity in tie strengths, and the frequent repetition of connections that it entails, allows in fact less virulent SIS-like diseases to survive in tightly connected local clusters that serve as reservoir for the virus. We validate this picture by evaluating the threshold of both processes in a real temporal network. Our findings confirm the important role played by non-Markovian network dynamics on dynamical processes.

PACS numbers: 89.75.Hc, 64.60.aq

Virtually any system can be represented as a network whose basic units are described as nodes and its interactions as links between them. In general, the connections are not static, but evolve in time subject to non-trivial dynamics. Consider for example face to face or online interactions networks where individuals talk and exchange informations through evolving contacts. The recent advances in technology have allowed to collect, monitor and probe such interactions generating an unprecedented amount of time-resolved high resolution datasets. The analysis of such real systems exposed the limits of canonical static and annealed networks representations calling for the development of a new theory to model, understand and account for network’s temporal properties. In particular, the recent data deluge has allowed to start identifying the effects that time varying topologies have on dynamical processes taking place in their fabric. Prototypical examples are the spreading of memes, ideas, and infectious diseases. All these phenomena can be successfully described as diffusion processes on contact networks and are affected by the ordering, concurrence, duration, and heterogeneity in nodes’ activities and connectivity patterns.

One of most distinctive property of social networks is the heterogeneity of interactions strengths. Individuals have memory of their inner circle of friends and most important connections activating some links more often then others, thus building up strong and weak ties with their peers. In other words, links creation is not a Markovian process. While this property has been observed, studied, and modeled in details in static networks, its understanding in the context of time-varying graphs is still far to be complete. Indeed, few studies have tackled this subject with different approaches uncovering a rich phenomenology. In particular, the non-Markovian link dynamics has been proven to be responsible either for slowing down/inhibit or, more surprisingly, speed-up/facilitate diffusion processes.

Here we study the effects of memory on two different classes of epidemic spreading models, namely the Susceptible-Infected-Recovered (SIR) and the Susceptible-Infected-Susceptible (SIS) models. We consider a recently proposed class of time-varying networks called activity driven models based on a simple reinforcement mechanism that allows reproducing many real properties of time-varying networks as the memory of individuals. Recently, this limitation has been overcome with the introduction of a non-Markovian generalization of the modeling framework based on a macroscopic fraction of the population. We study the dynamical properties of SIR and SIS models in activity driven networks with and without memory. In particular, we focus on one of the most important dynamical property of epidemic diffusion process, namely the epidemic threshold, defining the conditions necessary for the spreading of the disease to a macroscopic fraction of the population.

We find that memory acts in opposite ways on SIR and SIS models. In SIR models the epidemic threshold is shifted to larger values, making the spreading of the disease more difficult, while in SIS it moves to smaller values. Thus, non-Markovian dynamics facilitate the spreading of SIS-like diseases, like sexual transmitted illnesses, that can survive reaching an endemic state, reservoir, in tightly connected clusters. This difference between the two models is due to the fundamentally different nature the two processes that induce distinct behav-
iors also in the case of static networks \[51\]–\[53\]. In order to confirm the picture emerging in activity-driven networks, we consider also the real-world network built using messages exchanged between users on Twitter. Here each user is described as node and a link between two nodes is created if they interacted at least one time via message. Remarkably, the results are qualitatively similar to what observed in synthetic networks.

The paper is structured as follow. In section I we introduced activity driven networks without and with memory. In section II we study numerically the threshold behavior of SIR and SIS epidemic models in activity driven networks. In section III we analyze numerically the threshold of the two processes in the Twitter dataset. In section IV we draw our final remarks and conclusions.

I. ACTIVITY DRIVEN MODELS

In this section we describe the modeling framework used to produce the synthetic time-varying networks considered.

A. Memoryless activity driven models (ML)

In their basic formulation activity driven models are Markovian and memoryless. Each node is characterized by an activity rate \(a\), extracted from a distribution \(F(a)\), describing its probability per unit time to establish links. To account for the observation that human behaviors are characterized by broad activity distributions we will consider power-law distributions of activity \(F(a) = Ba^{a-\gamma}(\epsilon \leq a \leq 1)\), if not specified differently. In this setting, the generative process of the network is defined according to the following rules (see Figure 1):

- At each discrete time step \(t\) the network \(G_t\) starts with \(N\) disconnected vertices;
- With probability \(a_i\Delta t\) each vertex \(i\) becomes active and generates \(m\) links that are connected to \(m\) other randomly selected vertices. Non-active nodes can still receive connections from other active vertices;
- At the next time step \(t+\Delta t\), all the edges in the network \(G_t\) are deleted.

Thus, all interactions have a constant duration \(\Delta t\), that without loss of generality we fix to one, i.e. \(\Delta t = 1\).

At each time step the network \(G_t\) is a simple random graph with low average connectivity. Indeed, on average the number of active nodes per time step is \(N\langle a \rangle\), corresponding to an average number of edges equal to \(mN\langle a \rangle\), and an average degree \(\langle k \rangle = 2m\langle a \rangle\). However, integrating the links over \(T\) time steps, so that \(T/N \ll 1\), induces networks whose degree distribution follows the activity functional form \[29\]–\[33\] so that, for example, broad distributions of activity will generate broad degree distributions. The creation of hubs (highly connected nodes) results from the presence of nodes with high activity rate, which are more prone to repeatedly engage in interactions.

B. Activity driven models with memory (WM)

It has long been acknowledge that links in real-world networks can be grouped in (at least) two classes, namely strong and weak ties [37, 38]. The first represent connections that are activated often and describe the inner social circle of each node. The latter describe occasional contacts that are activated sporadically. The reason for these different classes of links is that individuals are non-Markovian, and the evolution of their ego-centered networks is deeply influenced by their memory. Interestingly, empirical observations indicate that the probability for an individual that had interacted with \(k\) people to initiate a connection towards a \(k+1\)th individual is a function of \(k\). More precisely, the analysis of a large-scale
mobile phone dataset [26] identified the relation

$$P(k + 1) = \frac{c_k}{k + c_k},$$  \hspace{1cm} (1)

where \(c_k\) is a constant function of the number of nodes already contacted, the degree. Thus, setting for simplicity \(c_k = 1 \forall k\), it is possible to generalize the activity driven framework accounting for individuals’ memory [26]. Given, as for the ML case, \(N\) nodes each characterized by an activity rate \(a\) extracted from a distribution \(F(a)\), the generative process of the WM network is defined according to the following rules (see Figure 1):

- At each discrete time step \(t\) the network \(G_t\) starts with \(N\) disconnected vertices;
- With probability \(a_i \Delta t\) each vertex \(i\) becomes active and generates \(m\) links;
- Each link is established with probability \(1/(k_i + 1)\) at random, and with probability \(k_i/(k_i + 1)\) towards one of the \(k_i\) previously connected nodes. Non-active nodes can still receive connections from other active vertices;
- At the next time step \(t + \Delta t\), the memory of each node is updated and all the edges in the network \(G_t\) are deleted.

The structural properties of time-aggregated ML and WM activity driven networks are fundamentally different. As clear from Figure 2-ML networks show a heavy-tailed cumulative degree and a homogeneous weight distribution, where the weights measuring the number of times each link is activated reflect the Markovian links’ creation dynamics (see 2-B). On the other hand, WM networks show a broad degree distribution, steeper than one observed in ML systems, (see Figure 2-A) and a heavy-tailed weight distribution indicating the heterogeneity of tie strengths (see Figure 2-B). In Figure 2-C we also compare the behavior of the largest connected component (LCC) integrating the links as function of time. Interestingly, in ML networks the LCC appears earlier. Memory slows down the growth of the connected component as individuals are more likely to activate previous connections.

II. SIR AND SIS MODELS IN ACTIVITY DRIVEN NETWORKS

We consider two classic epidemic models, namely the SIR and SIS model [50]. In both cases the population is divided in compartments indicating the health status of individuals. In the SIR model nodes can be in the susceptible (S), infected (I) or recovered (R) compartments. Susceptible nodes are healthy individuals. Infected nodes instead contracted the illness and can spread it. Recovered are nodes that healed from the disease becoming immune. The model is described by the following reaction scheme:

$$S + I \xrightarrow{\beta} 2I, \hspace{0.5cm} I \xrightarrow{\mu} R.$$  \hspace{1cm} (2)

The first transition indicates the contagion process. Susceptible nodes in contact with infected individuals become infected with rate \(\beta\). In particular, \(\beta\) takes into account the average contacts per node \(\langle k \rangle\) and the per contact probability of transmission \(\lambda\), i.e. \(\beta = \lambda \langle k \rangle\). The second transition instead, describes the recovery process. Infected individuals recover permanently with rate \(\mu\).

Whether the disease is able to spread affecting a macroscopic fraction of the network or not depends on the value of infection rate, recovery rate and networks dynamics.
In particular, in ML networks the SIR contagion process is able to spread if
\[ \frac{\beta}{\mu} \geq \xi^{SIR} = \frac{2 \langle a \rangle}{\langle a \rangle + \sqrt{\langle a^2 \rangle}}. \] (3)

See Refs. [29, 54] for the derivation details. The quantity \(\xi^{SIR}\) defines the epidemic threshold of the process. For value of \(\beta/\mu < \xi^{SIR}\) instead the disease will die out. Interestingly, the threshold is function of the first and second moments of the activity distribution, and completely neglects any time-integrated network representation.

In the SIS model nodes can be either in the susceptible (S) or infected (I) compartment. The model is described by the following reaction scheme:
\[ S + I \xrightarrow{\beta} 2I, \quad I \xrightarrow{\mu} S. \] (4)

The first transition is the same of SIR models. In the second transition instead infected individuals heal spontaneously but instead of becoming immune to the disease move back to the susceptible compartment with rate \(\mu\). In ML networks the epidemic threshold of SIS contagion process, \(\xi^{SIS}\), is:
\[ \frac{\beta}{\mu} \geq \xi^{SIS} = \frac{2 \langle a \rangle}{\langle a \rangle + \sqrt{\langle a^2 \rangle}}. \] (5)

See Refs. [29, 54]. Interestingly, the threshold is the same also for the SIR model, i.e. \(\xi^{SIS} = \xi^{SIR}\). This is a characteristic of ML activity driven networks and is due to the Markovian links’ creation dynamics [29, 32, 54].

In this paper we investigate numerically the epidemic dynamics occurring on WM networks.

**The SIR process on ML and WM networks**

We consider a SIR model and start the epidemic setting at \(t = 0\) with a fraction \(I_0 = 10^{-2}\) randomly selected nodes as seeds. SIR models reach the so called disease-free equilibrium in which the population is divided in:
\[ S_\infty + R_\infty = 1, \quad I_\infty = 0. \] (6)

All the variables refer to the density of individuals in the population. The infected individuals will always disappear from the population, as each one of them will eventually recovered becoming immune. Below the threshold, in the thermodynamic limit, \(R_\infty \to 0\). Above the threshold instead \(R_\infty\) reaches a macroscopic value, i.e. \(R_\infty = \mathcal{O}(1)\). The transition between the two regimes is continuous and the behavior of \(R_\infty\) can be studied as a second order phase transition with control parameter \(\beta/\mu\) [2, 3].

In Figure 3 we show the results obtained measuring \(R_\infty\) in ML and WM networks for different values of \(\beta/\mu\). Without loss of generality we fix \(\mu = 10^{-2}\) and use \(\beta\) as free parameter. The epidemic threshold in WN networks is clearly larger than in ML systems. The memory of individuals shifts the threshold to larger values, making the systems less vulnerable to disease spreading, as the repetition of interactions within strong ties inhibits the spreading potential of the disease. Indeed, infected individuals will contact with higher probability their inner circle of ties infecting possibly some of them. However, the newly infected nodes will be prone to keep contacting back the initial seeds and eventually recover. On the contrary, in ML networks nodes initiate random connections at each time step increasing their probability of interacting with infected individuals. It is also interesting to notice that in WM networks the final fraction of individuals affected by the disease is smaller. For example, for values of \(\beta/\mu = 1\), \(R_\infty\) reaches a fraction 0.35 in ML networks, while just 0.15 in WM graphs. In summary, memory roughly doubles the epidemic threshold of a SIR process making the system more resilient to the spreading, and above threshold reduces the size of affected population.

**The SIS process on ML and WM networks**

We now turn our attention to SIS processes. Also in this case we start the epidemic setting at \(t = 0\) a fraction \(I_0 = 10^{-2}\) of randomly selected nodes as seeds. The nature of this epidemic model is fundamentally different respect to the SIR. Indeed, above threshold SIS processes show an endemic state characterized by a constant fraction of nodes, \(I_\infty > 0\), in the infected compartment. Below threshold instead, the process reaches a disease-free equilibrium, i.e. \(I_\infty = 0\). In order to estimate the critical value of \(\beta/\mu\) we proceed as seen for the SIR model, considering in this case the behavior of \(I_\infty\). In Figure 4-A we show the results of SIS processes evolving on top of ML and WM activity-driven networks. While the threshold of SIS coincides with the one of SIR in ML networks, it becomes smaller in WM networks.
FIG. 4: SIS model. A) $I_{\infty}$ as a function of $\beta/\mu$ in ML (blue circles) and WM (red squares) activity driven networks. B) Average lifetime $L$ of SIS processes as a function of $\beta/\mu$ in ML and WM activity driven networks. The peak and vertical lines show the estimated threshold. C) Survival probability $P_s(t, \beta/\mu)$ of SIS processes as a function of $t$ in ML and WM activity driven networks for two different values $\beta/\mu$. For all the panels we set $N = 10^5$, $\epsilon = 10^{-3}$, $m = 1$, and $\mu = 10^{-2}$. Each point is the evaluated considering $10^2$ independent simulations started with a fraction of $10^{-2}$ randomly selected seeds.

In general, in SIS processes the numerical estimation of the threshold is more prone to size and noise effects, due to the subtleties relative to the identification of the endemic state and the fact that $I_{\infty}$ is not a monotonically increasing quantity as in the case of $R_{\infty}$. Therefore, we consider also the life time $L$ and the coverage $C$ of the process as a function of $\beta/\mu$, defining the duration of the process and the fraction of nodes that acquire the infection, respectively. In SIS processes for values of $\beta/\mu$ above threshold the life time is infinite (endemic state) and the coverage reaches 1. Instead, below threshold both $L$ and $C$ vanish in the thermodynamic limit. Interestingly, the life time obtained by averaging over many realizations is equivalent to the susceptibility $\chi$ in standard percolation theory. This method allows us to detect the threshold precisely [55]. Indeed, following Ref. [54] we can consider as above threshold any realization that reaches a macroscopic coverage $C$. Without loss of generality we set $C = 0.5$, fix $\mu = 10^{-2}$ and evaluate $L$ as a function of $\beta$. For small values of the contagion rate the disease dies out quickly and the coverage remain below the threshold $C$, while for very large values of $\beta$ the disease will be able to spread quickly reaching a fraction $C$. For intermediate values of $\beta$, $L$ will increase showing a peak close to the actual epidemic threshold. Figure 4B shows that the estimation of the threshold performed considering the life time of the process produces results very similar to what observed in Figure 4A. We can therefore conclude that the threshold of a SIS process unfolding in WM networks is smaller than in ML systems. This behavior is quite surprising and opposite to what is observed in the case of SIR models. The repeated connections in the ego-centered networks of each node allow the disease to survive in local and small clusters of strong ties making the system more fragile to the disease spreading. Such a behavior is not observed in SIR processes due to the presence of recovered individuals that become immune to the disease and are unable to sustain the spreading with multiple reinfections.

In order to further substantiate the picture emerging from these analysis we study the surviving probability, $P_s(t, \beta/\mu)$, of SIS processes in the two networks [50], defined as the probability that the process survives, $I(t) > 0$, up to time $t$. Above threshold the disease reaches an endemic state implying a constant $P_s(t, \beta/\mu)$. Below threshold instead $P_s(t, \beta/\mu)$ decays to zero [57]. In Figure 4C we show the results in ML and WM activity driven networks considering for both cases values of $\beta/\mu$ above and below the threshold estimated in Figure 4B. The survival probability is constant and equal to one for the two values of $\beta/\mu$ estimated to be above threshold [58]. Instead, it decays for the other two values below the threshold. This analysis confirms that the threshold of SIS processes in activity driven networks with memory is smaller than the threshold in memoryless networks.

III. SIS AND SIR MODELS IN REAL TIME VARYING NETWORKS

In order to validate the results obtained on synthetic time-varying networks we study the dynamical properties of SIR and SIS processes on a real temporal system. We consider the interactions between 117664 Twitter users in the first tree months of 2008 via 359226 messages and coarse-grain the data adopting a time resolution of one day. Each user is represented as a node, and an undirected link is drawn between two nodes if they exchanged at least one message. Arguably such network is driven by non-Markovian human dynamics as many users tend to interact several times with the same circle of accounts.
In order to single out the effects of memory we consider also a randomized version of the considered Twitter network, where non-Markovian dynamics are eliminated. The randomization is performed by reshuffling the interactions for each time stamp, so that memory effects are removed while the sequence of activation times for each node, the final time integrated degree distribution, and the degree distribution at each time step are preserved as a function of $\beta/\mu$ in the real (red squares) and in the randomize (blue circles) Twitter networks. B) SIS dynamics. Life time $L$ in the real (red squares) and in the randomize (blue circles) Twitter networks. Each point is evaluated considering $10^5$ independent simulations started with a fraction of $10^{-2}$ randomly selected seeds. We set $\mu = 0.5$ in both cases.

In conclusion, the results here presented show that memory acts on SIR processes unfolding on realistic networks, where non-Markovian dynamics are eliminated. The randomization is performed by reshuffling the interactions for each time stamp, so that memory effects are removed while the sequence of activation times for each node, the final time integrated degree distribution, and the degree distribution at each time step are preserved as a function of $\beta/\mu$ in the real (red squares) and in the randomize (blue circles) Twitter networks. B) SIS dynamics. Life time $L$ in the real (red squares) and in the randomize (blue circles) Twitter networks. Each point is evaluated considering $10^5$ independent simulations started with a fraction of $10^{-2}$ randomly selected seeds. We set $\mu = 0.5$ in both cases.

IV. CONCLUSIONS

In general, real networks are characterized by temporal and non-Markovian dynamics. For example, in social networks, individuals interact more frequently with a small set of strong ties. In other words, people keep memory of their past connections. While this crucial aspect has been analyzed in details in static networks’ representations, very little attention has been devoted to its characterization on temporal networks. Here we studied the dynamical properties of SIR and SIS models in activity driven networks with and without memory. In order to single out the effects of non-Markovian dynamics we studied the epidemic threshold in basic activity driven models that by construction are Markovian and memoryless, and in a recent generalization of this modeling framework that explicitly consider non-Markovian link dynamics. We found that memory acts on SIR processes shifting the epidemic threshold to larger values making the system more resilient to the disease spreading. On the contrary, memory acts on SIS processes by lowering the epidemic threshold to smaller values thus making the systems more prone to the disease invasion. In fact, the heterogeneity in ties’ strength induces frequent repetition of contacts that allow the survival of SIS-like diseases in local groups of tightly connected individuals. The illness reaches its endemic state in small clusters that act as reservoir for the virus.

Although activity driven models with memory capture fundamental aspects of real time varying networks, they do not account for other important features as appearance of new nodes, disappearance of old ones, and bursty behaviors just to name a few. While the introduction of these ingredients is left for future work, here we validated the picture obtained from synthetic networks by considering a real time-varying system, namely the network of communications in Twitter. Interestingly, the results obtained in this case confirm qualitatively the findings observed in activity driven networks.

In conclusion, the results here presented show that memory can have opposite effects on different classes of spreading processes, and corroborate the important role played by non-Markovian dynamics on the dynamical processes unfolding on realistic networks.

V. ACKNOWLEDGMENTS

The authors are grateful to Alessandro Vespignani for helpful discussions, insights, and comments.
J. F. F. Mendes, Physical Review Letters 109, 128702 (2012).
[54] M. Starnini and R. Pastor-Satorras, Phys. Rev. E 89, 032807 (2014).
[55] M. Boguña, C. Castellano, and R. Pastor-Satorras, Phys. Rev. Lett. 111, 068701 (2013).
[56] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. 86, 3200 (2001).
[57] M. Muñoz, R. Juhász, C. Castellano, and G. Ódor, Phys. Rev. Lett. 105, 128701 (2010).
[58] This observation is important to exclude the presence of Griffiths phases.