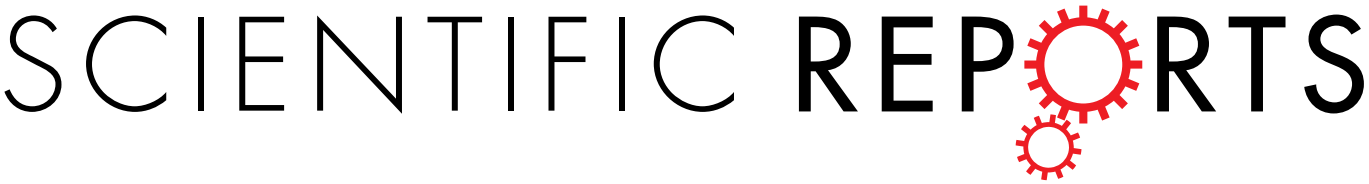
Epidemic spreading in modular time-varying networks



OPEN

Epidemic spreading in modular time-varying networks

Matthieu Nadini[1,6], Kaiyuan Sun[2], Enrico Ubaldi[3], Michele Starnini[4,5], Alessandro Rizzo[6] & Nicola Perra[7]

Received: 27 October 2017
Accepted: 17 January 2018
Published online: 05 February 2018

We investigate the effects of modular and temporal connectivity patterns on epidemic spreading. To this end, we introduce and analytically characterise a model of time-varying networks with tunable modularity. Within this framework, we study the epidemic size of Susceptible-Infected-Recovered, SIR, models and the epidemic threshold of Susceptible-Infected-Susceptible, SIS, models. Interestingly, we find that while the presence of tightly connected clusters inhibits SIR processes, it speeds up SIS phenomena. In this case, we observe that modular structures induce a reduction of the threshold with respect to time-varying networks without communities. We confirm the theoretical results by means of extensive numerical simulations both on synthetic graphs as well as on a real modular and temporal network.

Network thinking has become a prominent and convenient paradigm to unveil the properties of complex systems. Such a paradigm has been rapidly enriched, giving rise to variants that account for inherent features of real complex systems inferred by the availability of large, often time-resolved datasets[1–4]. Three are the main features that have captured the attention of researchers in the area. The first is heterogeneity in the statistical distributions of key topological properties such as the number of connections per node (degree) and the intensity of interactions (weight)[3,4]. This property is one of the hallmarks of complexity and is linked to a range of non-trivial dynamics[3,5]. For example, heterogeneity in the connectivity patterns makes networks extremely fragile to the spreading of infectious diseases and malicious attacks[6,7]. The second feature regards the presence of modules and communities[8]. Available datasets have highlighted that real networks are organized in modules, or communities, whereby the density of links *within* the community is much greater than the density of links *between* communities. On the one hand, communities can be treated as fairly independent entities within a large network, like the behaviour of different organs within the same body. On the other hand, from time to time, phenomena originating in a community may involve a huge portion, if not all, of the network. This is for example the case of pandemics originating from local outbreaks[6,9–11]. Nevertheless, the role played communities is still ambivalent. For example, the presence of communities might slow down or speed up the propagation of a disease and facilitate the spreading of social norms[12–20]. Detecting communities in a real system is not a trivial task, due to their fuzzy, vanishing, and overlapping nature. Also, in most of the available datasets communities are not explicitly labeled, so the validation of community detection algorithms cannot often be rigorously carried out[8,21]. Finally, networks are characterised by non trivial temporal dynamics[22,23]. The propensity of nodes to initiate and to attract interactions per unit time is typically heterogeneously distributed[24,25]. The same applies to the duration and the time interval between connections[13,26]. Furthermore, the creation or renewal of interactions might be correlated, and the dynamics driving the temporal evolution of networks are function of the time-scale considered[27,28]. However, the large majority of studies on dynamical processes unfolding on networks have been conducted under the hypothesis of *time scale separation* which effectively neglects all such features. In particular, the evolution of the process and the evolution of network are considered to take place at well-distinct time scales. Within this paradigm two opposite limits have been considered. In the first case, the dynamical process is assumed to be much faster than the evolution of the network. This is the limit of quenched/static networks[29]. Here, networks are fully characterised from an adjacency

[1]Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn, NY, 11201, USA. [2]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, 02115, USA. [3]Institute for Scientific Interchange, ISI Foundation, Turin, Italy. [4]Departament de Física Fondamental, Universitat de Barcelona, Martí i Franquès 1, 08028, Barcelona, Spain. [5]Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain. [6]Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy. [7]Centre for Business Networks Analysis, University of Greenwich, London, UK. Correspondence and requests for materials should be addressed to N.P. (email: n.perra@greenwich.ac.uk)

matrix, A_{ij} whose entries are non-zero for all connected pairs $i-j$[3]. The second case instead, is the opposite limit where an annealed version of the network can be considered and averaging (mean-field) techniques can be applied[30–32]. Annealed networks are fully characterised by an average adjacency matrix $\overline{A}_{k_i,k_j}$ describing the probability of connection for nodes of degree k_i and degree k_j[33]. In our case instead, the time scale regulating the dynamical process and the evolution of the network are comparable, the time scale approximation is not valid. This is the regime of time-varying networks[23]. Interestingly, the temporal nature of interactions might inhibit or facilitate spreading processes evolving at comparable time-scales[11,24,34–46]. The effects introduced by communities and time-varying connectivity patterns on dynamical processes have been mostly scrutinised separately. However, as few recent works pointed out, the two attributes are connected and their interplay introduces non-trivial effects, such as segregated behaviours and formation of hierarchical structures[47]; or the intricate competition between topological and temporal correlations[48]. The presence of groups, think for example the interaction network of students in a school, introduces specific dynamics that deeply affect spreading processes. A thorough modelling and study of these phenomena may be useful to define prioritisation of interventions and containment strategies in epidemic spreading[49].

Altogether, these observations call for a general modelling framework aimed at characterising both features and single out their effects on real networks. The model presented in this paper leverages on the paradigm of Activity Driven Networks (ADNs) to model realistic temporal networks where the node and link dynamics coevolve at comparable time scales[24,50] and includes the modularity phenomena, whereby connection patterns can be set to preferentially occur within a given community, rather than outside the community toward the rest of the population. While the coevolution of node and link dynamics is inherently considered in ADNs by construction, modularity is here modelled by a single parameter that regulates the interplay between the link formation within and outside of the community. In the context of epidemic processes on time-varying networks, our model is first characterised analytically. Then it is used to study the behaviour of different contagion processes on synthetic networks, and on a large, time-resolved dataset of scientific collaborations. Results and methods are discussed in detail in the following sections.

Results

Here, we study the effect of modularity (i.e., the presence of communities in the network) on time-varying networks. To this extent, we introduce a model of time-varying networks with tunable modularity, able to capture several features of real temporal graphs. We derive an analytical characterisation of the model, and we study the behaviour of the Susceptible-Infected-Recovered (SIR) and the Susceptible-Infected-Susceptible (SIS) epidemic processes unfolding on its fabric[51]. Remarkably, while the presence of tightly connected clusters inhibits SIR processes, it favours the spreading of SIS-like diseases lowering the epidemic threshold. Interestingly, similar results have been recently obtained in models of time-varying networks characterised by correlated topological features induced by reinforcement of specific ties[42]. We confirm the theoretical picture emerging from synthetic networks by means of extensive simulations on a real word dataset of scientific collaborations within the American Physical Society (APS). Our results contribute to characterise the mechanisms, and their interplay, behind the complex, and often contradictory, behaviour of dynamical processes unfolding on real networks.

Modular activity driven networks. The system under investigation is composed by N nodes. Each node i is characterised by an activity rate a_i, that describes its propensity to engage in social interactions with other nodes. To capture empirical observations performed in a wide set of systems ranging from R&D to online interactions networks[25,38,52,53], we consider activity rates heterogeneously distributed, and extracted from a continuous functional form $F(a) = Ba^{-\nu}$, where $a \in [\varepsilon, 1]$ and $\varepsilon = 10^{-3}$ to avoid divergence in the distribution. Furthermore, each node is assigned to only one group/community. To take into account empirical evidences, the size of each community is extracted from a heavy-tailed distribution, i.e. $P(s) = Cs^{-\omega}$ with $s \in [s_{min}, \sqrt{N}]$[8,54]. Therefore, we do not limit ourselves in studying a fixed number of modules[55], whilst their number is driven from the model's parameters. The assignment of nodes to communities is done as follow. A community of size s is extracted. The ID of s nodes is progressively assigned to the community. The processes is repeated until all N nodes are assigned to one community. Very rarely, all nodes can be perfectly assigned to the extracted community sizes. Indeed, in the last extraction we might have available only a fraction of the nodes necessary to fill the community. However, the average value of s is much smaller than the total number of nodes, thus the actual size of the last community can be only slightly smaller than the extracted value. The empirical distribution of community sizes in the network will then follow $P(s)$. Given these settings, a generative network model is defined by the following steps (see Fig. 1).

- At each time t, the network, G_t, starts with N disconnected nodes.
- With probability $a_i\Delta t$ each vertex i is active and willing to create m connections.
- Each link being generated points with probability μ within the node's community, and with probability $1-\mu$ to one of any other groups. In both cases, the target node j of the link is randomly selected in the target community.
- At the next time step $t+\Delta t$ all the edges in G_t are deleted.

All the interactions have a constant duration Δt. In the model, neither self-loops nor multiple edges are allowed. In the following, without loss of generality, we fix $\Delta t = 1$. Furthermore, we consider the case $m = 1$.

Given an heterogeneous distribution of activity, at each time step, the model generates a random, structureless network in which few nodes are active. The modular features of the network emerge integrating connections in time. Such time-integrated properties, at different time regimes, can be computed analytically. In the following,

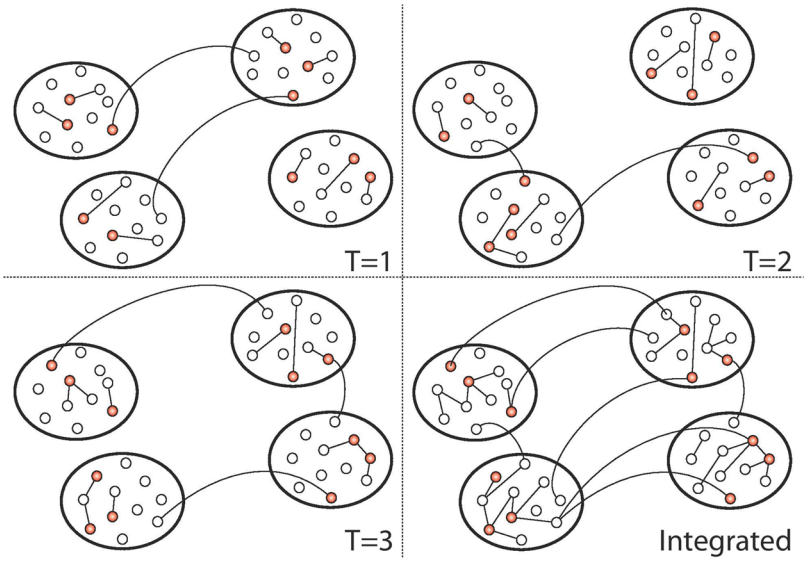


Figure 1. Schematic representation of the model. In red, we show active nodes. Straight lines and arcs describe links connecting nodes in the same or in different communities respectively. In the bottom right panel we show the integrated network obtained as the union of G_1, G_2, G_3.

we will report the results for the evolution of the average number of connections of each node $\langle k_i(t)\rangle$ (average degree) and the overall degree distribution $\rho(k)$. The complete set of results is shown in the Supplementary Information (SI).

To solve the average degree's dynamics, let us introduce the effective activity $\tilde{a}_i = a_i + \langle a\rangle$ (where $\langle a\rangle$ is the average value of the activity distribution) and the mixing parameter $\mu' = 1-\mu$. We refer to the degree of node i at time t as $k(a_i, s, t)$, where s is the node's community size. By defining an activity class as the group of nodes featuring similar activity values a, we set the average in-community degree $\langle k_c(a, s, t)\rangle$ to be the average number of connections that nodes belonging to the activity class a and falling in communities of size s have toward nodes of their same community. The latter grows as

$$\langle k_c(a, s, t)\rangle = (s-1)\left[1 - \exp\left(-\frac{t}{\tau(a, s)}\right)\right], \tag{1}$$

where $\tau(a, s)$ is the characteristic time that it takes for the degree $k_c(a, s, t)$ of nodes of activity a belonging to a community of size s to be $k_c(a, s, t) \sim (s-1)$, being $s-1$ the maximum value of the in-community degree (see the Supplementary Information for the evaluation of $\tau(a, s)$).

Similarly, we can define the average out-community degree $\langle k_o(a, t)\rangle$ as the number of connections that nodes of activity class a have outside of their communities at time t. We expect this quantity to be independent of the nodes' community size s so that, for large networks, we can write:

$$\langle k_o(a, t)\rangle = \mu'\tilde{a}t \tag{2}$$

The average total degree $\langle k(a, s, t)\rangle$ can be computed as the simple sum between the two previous equations, obtaining

$$\langle k(a, s, t)\rangle = \langle k_c(a, s, t)\rangle + \langle k_o(a, t)\rangle \simeq \begin{cases} \tilde{a}t & t \ll \tau(a, s) \quad (3a)\\ \mu'\tilde{a}t + (s-1) & t \sim \tau(a, s) \quad (3b)\\ \mu'\tilde{a}t & t \gg \tau(a, s) \quad (3c)\end{cases}$$

Three regimes are readily identified: an initial growth in which both the in-community and the out-community degrees are growing linearly in time, followed by the slowing down of the in-community degree, which saturates to $s-1$, and then a further linear regime driven only by the out-community degree growth. Figure 2 shows that the numerical simulations perfectly match with the theoretical formulas (see the SI for details).

Noticeably, the long time evolution of the node degree is linear in time and proportional to its activity class a, so that we find the asymptotic degree distribution of the system to feature the same functional form of $F(a) \propto a^{-\nu}$, as found in non-modular activity driven networks[24,56]:

$$F(a)da \xrightarrow{k(a,t)\propto a\cdot t} \rho(k)dk \propto k^{-\nu}dk. \tag{4}$$

In Fig. 3, we integrate the network for $T = 10^5$ and we plot the three degree distributions. As expected, the out-community $\rho(k_o)$ and the total $\rho(k)$ degree distributions fall as power laws with exponent $-\nu$. On the other

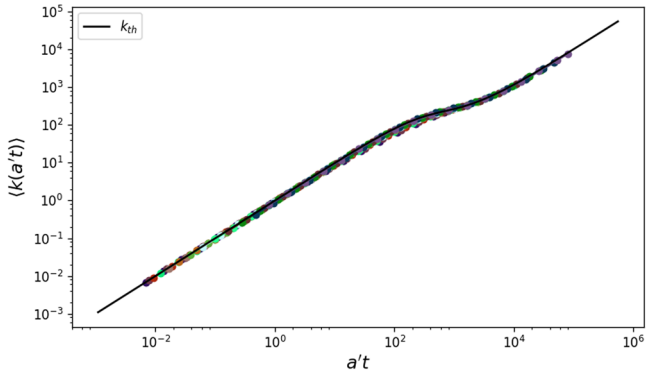


Figure 2. Time evolution of the average total degree, $\langle k(a, s, t)\rangle$, for different activity classes and compared with the theoretical function of Eq. 3a,b,c and, evaluated considering a community size equal to the average (i.e. $s = \langle s\rangle$). The rescaled time is $t \to \tilde{a}t$ and $\langle k(\tilde{a}t)\rangle$ is plotted. Parameters used are: $N = 10^5$, $\omega = 2.1$, $\nu = 2.1$, $s_{min} = 10$, $\mu = 0.9$ and $T = 10^5$ evolution steps. Each point is an average of 10^2 simulations.

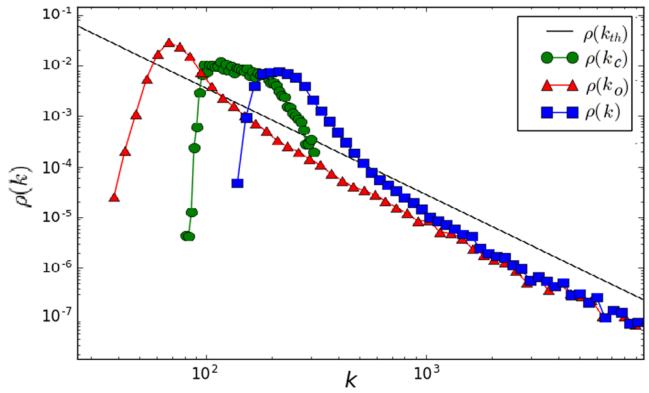


Figure 3. Plot of the three degree distributions and the theoretical prediction, given in Eq. 4. Parameters used are: $N = 10^5$, $\omega = 2.1$, $\nu = 2.1$, $s_{min} = 10$, $\mu = 0.9$ and $T = 10^5$ evolution steps.

hand, the in-community degree $\rho(k_c)$ saturates to the community size distribution $P(s)$, as all the nodes reach their maximum in-community degree value $(s - 1)$, being that the modules' size is far smaller than the network size ($s_{\max} = \sqrt{N} \ll N$). On the contrary, the out-community degree takes longer times to saturate to its maximum value $N - s \gg s$.

It is worth stressing that the results presented in this section apply to the networks obtained integrating links over time. A process unfolding on such networks, in general, will be affected by the time-aggregated features of the graph. The extent to which this is true, is function of the interplay between the time-scale describing its evolution, τ_P, and the various $\tau(a, s)$. In the limit $\tau_p \ll \tau(a, s)$ the process would effectively evolve on the instantaneous, annealed networks that are characterised by a small average degree and modularity. In the opposite limit instead, the process would effectively unfold on static networks obtained integrating links over longer time characterised by high average degree and low modularity. Indeed, the average degree in this regime will be dominated by out-community links that make the connections between different communities increasingly stronger, thus increasingly destroying the identity of communities. In the limit $\tau_p \sim \tau(a, s)$ the process would effectively evolve on maximally modular networks (for a given set of parameters). Arguably, this is the most interesting regime that we will consider in the following.

Epidemic processes on modular activity driven networks. Let us turn our attention on the dynamical properties of SIR and SIS processes (see the Methods section for a detailed definition of the two) unfolding on the proposed model. Although similar, the two processes are intrinsically different[33,57–59]. Indeed, SIR processes are always characterised by the so called disease-free equilibrium, provided $d_t N = 0$. The illness eventually disappear, i.e., $I = 0$ for $t \to \infty$. SIS models instead allow the existence of an endemic state where a finite and constant fraction of infected individuals permanently colonise the population, i.e., $I > 0$ for $t \to \infty$. Here, we focus on a central concept of contagion phenomena: the epidemic threshold. This quantity defines the conditions necessary for the spreading of the illness. In annealed networks, the threshold is determined by the moments of the degree distribution $\rho(k)$. In static graphs the expression is given by the principle eigenvalue of the adjacency matrix[57,60,61]. In time-varying networks instead, the threshold is determined by the interplay between the time-scales of the contagion and network evolution processes[24,41,45,50,62–69]. In the case of SIR models, we also consider another important

quantity: the epidemic size R_∞ which is defined as the final ratio of recovered nodes. This describes the fraction of nodes affected by the disease.

To develop a deeper understanding, let us derive the mean-field level dynamical equations describing the contagion process in modular activity driven networks. We define the activity block variables $S_{a,s}$, $I_{a,s}$, and $R_{a,s}$ as the number of susceptible, infected and recovered individuals, respectively, in the class of activity a and community of size s at time t (to enhance readability, we omit to notate the dependence on time). This allows us to write the mean-field evolution of the number of infected individuals, for a SIR process, in each group of nodes with activity a as:

$$\begin{aligned} d_t I_{a,s} &= -\gamma I_{a,s} + \lambda S_{a,s}\left[\mu a\frac{I_s}{s} + (1-\mu)a\frac{I}{N}\right] \\ &\quad + \lambda\sum_{a'} a'\left[\mu I_{a',s}\frac{S_{a,s}}{s} + (1-\mu)I_{a',s}\frac{S_{a,s}}{N}\right], \end{aligned} \tag{5}$$

where I_s and I are the number of infected in communities of size s and in the whole network, respectively. The first term in the r.h.s. accounts for the recovery of infected individuals. The other four terms account for the probability that a Susceptible node in a community of size s connects to an Infected node inside (first) or outside (second) its community acquiring the infection, and for the probability that an Infected node of class a' connects to a Susceptible node inside (third) or outside (forth) a community of size s, contracting the disease. For simplicity, we consider that $N-s\sim N$ and, at least initially, $I-I_s\sim I$. Summing over all the activities and community sizes, and considering only the first order terms in a, $I_{a,s}$, $R_{a,s}$ and their products, we obtain

$$d_t I = -\gamma I + \lambda\langle a\rangle I + \lambda\Theta + \lambda\mu\sum_s(\langle a\rangle_s - \langle a\rangle)I_s, \tag{6}$$

$$\begin{aligned} d_t\Theta &= -\gamma\Theta + \lambda\langle a^2\rangle I + \lambda\langle a\rangle\Theta \\ &\quad + \lambda\mu\sum_s[(\langle a^2\rangle_s - \langle a^2\rangle)I_s + (\langle a\rangle_s - \langle a\rangle)\Theta_s], \end{aligned} \tag{7}$$

where we defined $\Theta = \sum_a aI_a$, and $\Theta_s = \sum_a aI_{a,s}$. The term $\langle a^x\rangle_s = \sum_a N_{a,s}a^x/s$ describes the moments of the activity distribution in any community of size s. The second, auxiliary, equation is obtained from the first by multiplying both sides by a and summing over all s and a. The epidemic threshold, in principle, can be derived evaluating the principle eigenvalue of the Jacobian matrix of the system of differential equations in I and Θ[24,41,45,50,69,70]. In general, a closed expression for the threshold does not exist. However, we can point out some interesting observations.

First of all, the terms associated to $R_{a,s}$ vanish, implying that, at the first order, the thresholds of both SIR and SIS are equal[50]. Furthermore, the terms in μ weight a comparison between the moments of the activity distribution in the network with the corresponding quantities evaluated inside each community. If fluctuations of these terms are negligible, due for example to very large community sizes or to narrow distribution of activity, the equations become equivalent to the case $\mu=0$. In the limit $\mu\to 0$ the network has no modular structure, and the threshold, for both SIR and SIS, becomes $\beta/\gamma \geq 2/(1+\sqrt{\chi})$ as derived with different approaches in refs[24,41,45,71]. We defined $\chi=\langle a^2\rangle/\langle a\rangle^2$, where the moments are evaluated over the whole network. As expected, the spreading condition is determined by the interplay between the time-scale of the contagion process and the time-scale of the network. Furthermore, the threshold is significantly larger with respect to the case in which the disease would spread in static or annealed networks generated integrating connections over time[24]. Indeed, the concurrency of contacts in the two time scale separation regimes drastically facilitates the spreading of diseases[23,72]. It is important to notice how the threshold is not function of m even in the case of $m>1$. This is due to the fact that we absorbed the contact rates in the definition of β. As detailed in the methods section, this is defined as the per capita rate of infection. By adopting such definition, we are able to estimate the spreading power of a disease independently of possible differences in contacts rates. The expression could be easily changed to explicitly account this aspect obtaining $\lambda/\gamma \geq \frac{1}{m\langle a\rangle(1+\sqrt{\chi})}$[24,46,73]. In the opposite limit $\mu\to 1$ networks are extremely modular, fluctuations become important and the symmetry between SIR and SIS breaks. In order to understand this limit, let us consider first a SIR process started from a single infected node in a community of size s. The large majority of connections are towards a small number of vertices in the same group. As soon as the disease start to spread within the community the number of infected and then recovered nodes grows, thus the probability of links $I-I$ and $I-R$ increases. Such connections cannot help the spreading of the disease. In fact they hamper the contagion process. In case instead of a network characterised by smaller values of modularity the growth of infected and recovered nodes has a much smaller effect. Indeed, the connectivity between communities would guarantee access to larger pool of susceptible nodes to sustain the spreading. From these simple observations we can expect that SIR processes are inhibited by highly modular connectivity patterns. Except for few exceptions in particular topologies[17], this is the case on static and annealed graphs[12–17]. As we will show below, the same arguments hold also in time-varying connectivity patterns. On the other hand, in case of SIS processes, the repetition of contacts does not lead to such "pair annihilation": contacts between infected nodes do not help the spreading of the disease, but they are only temporary (eventually, all infected nodes become susceptible again). Thus, we expect that modularity plays a different role in SIS dynamics. This is what has been found also in the case of one annealed network model[18]. Below we will show how this applies also in the case of time-varying networks. In order to numerically characterise SIR models, we study the epidemic size, R_∞, as a function of β/γ. This quantity

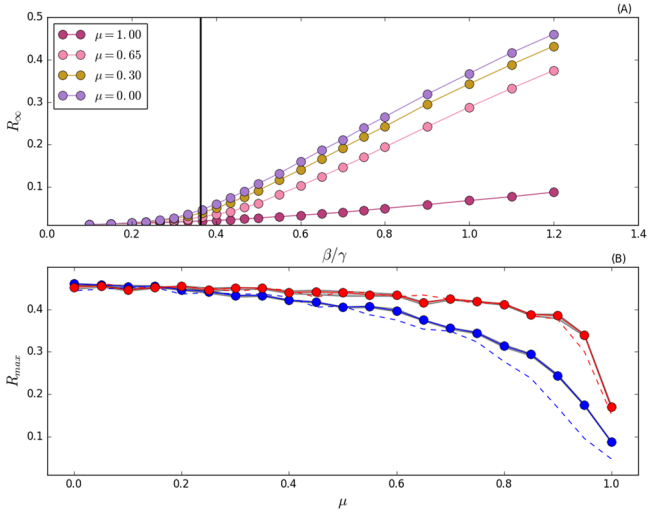


Figure 4. Panel (A) R_∞ as a function of β/γ, for selected values of μ and $s_{min} = 10$. Vertical black line represents the theoretical value of the epidemic threshold for $\mu = 0$ as derived in refs[24,71]. Panel (B) R_{max}, i.e. the max value of R_∞, as a function of μ. In red curves we set $s_{min} = 100$, in blue curves $s_{min} = 10$. In solid curves, we draw community sizes directly from the community size distribution $P(s)$. In dashed curves, we fix the community sizes as equal to the average value of $P(s)$ for all communities. The 95% confidence interval is in grey. Each point is an average of 10^2 independent simulations.

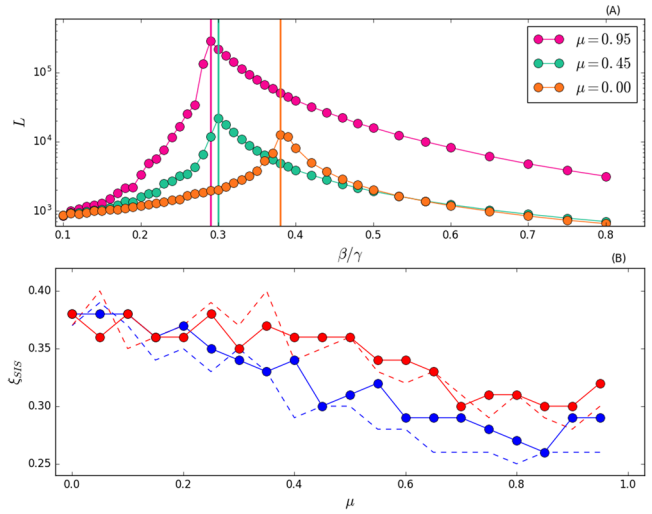


Figure 5. Panel (A) Lifetime of the disease L as a function of β/γ, for selected values of μ and when $s_{min} = 10$. Vertical lines are the epidemic threshold. Panel (B) Ratio $\xi_{SIS} = \beta/\gamma$ in correspondence of L_{max}, as a function of μ. In red curves we set $s_{min} = 100$, blue curves $s_{min} = 10$. Each point is an average of 10^2 independent simulations. Note that we avoid to simulate $\mu = 1$ because the criterion we follow for the estimation of the threshold does not hold for a network with many connected components.

acts as the order parameter of a second-order phase transition[6]. For SIS processes instead, the order parameter is the final fraction of infected individuals, I_∞[6]. The numerical estimation of this quantity is challenging, since it requires the precise determination of endemic states. For these reasons, we follow ref.[74], measuring the life time of the disease, L, that acts as the susceptibility in phase transitions[74,75]. This quantity is defined as the average time it takes for the disease to either die out or reach a macroscopic fraction, Y, of the populations. Without loss of generality, we start our simulations by setting 1% of randomly selected nodes as initial infected seed. Other parameters are set as: $\gamma = 0.01$, $m = 1$, $\nu = 2.1$, $\omega = 2.1$, $N = 10^5$ and $Y = 0.5$ (see SI for similar plots obtained fixing $\omega = 1.5$).

Results obtained from SIR models are represented in Fig. 4A,B, whilst results from SIS models are shown in Fig. 5A,B. In Figs 4B and 5B we study different community structure, either by considering a constant community size (dashed curves) or by drawing community sizes directly from the community size distribution $P(s)$ (solid curves). In general, red curves represents a network with bigger communities than the one represented with blue curves.

For SIR models, Fig. 4A tells us that, as expected, the higher β/γ the higher the epidemic size. The figure also confirm the intuitions about the threshold. Indeed, we observe a dependence on μ: the higher μ the higher the threshold. However, it is important to notice how such dependence is weak especially when compared with the SIS case (see below). Moreover, the higher the fraction of links created between pair of nodes sharing the same community (i.e. the higher μ), the lower the epidemic size. This second observation is confirmed studying different community structures, as done in Fig. 4B, in which we plot the maximum epidemic size (corresponding to the largest value of β/μ in our settings), R_{max}, as a function of μ. In the limit $\mu \to 0$, we observe that the disease impact is the same: the networks behave as if no community structure was present. Instead, when $\mu \to 1$, the modular structure influences the spread of the disease. As mentioned before, repeating contacts within communities significantly narrows the chances of having new infected individuals. Indeed, in SIR models, once a node recovers, it cannot be infected again. Repeating contacts with nodes already recovered does not favour the spread of the disease. Overall, the main observations are four. (i) Increasing the modularity reduces the epidemic size. (ii) A network with, on average, larger modules is likely to yield a higher epidemic size. (iii) The larger the modules the weaker the dependence on μ of the epidemic size. (iv) In case of small modules, the distribution of community sizes seems to influence the spreading of the disease. In particular, a network organised in small groups of constant sizes leads to smaller epidemic size respect to a network in which the average community size is the same, but individual sizes are extracted from a power-law distribution.

For SIS models, the lower μ, the lower the life time L (see Fig. 5). Inter-community links speed up the disease spreading and an endemic state, i.e. $Y = 0.5$, is reached faster. Moreover, the higher μ, the lower the epidemic threshold. This last observation, which implies that increasing values of modularity favour the survival of the disease, is confirmed in Fig. 5B where we also test the effects of different community structures. In the limit $\mu \to 0$, there is no community structure and the curves converge to the same epidemic threshold. On the contrary, when $\mu \to 1$, the community structure becomes increasingly important and influences the spreading. Qualitatively, higher levels of modularity diminish the epidemic threshold. This is due to the repetition of the same contacts within a community which becomes increasingly more likely. Indeed, in SIS models, reinfection is allowed and nodes can become infected many times: communities act as a reservoir for the disease and favour the contagion process pushing the epidemic threshold to smaller values. Besides this last point, there are two main observations. (i) A network with larger modules is likely to have an higher epidemic threshold. (ii) In case of communities with smaller average sizes and high values of modularity, having the community size extracted from a power-law seems to slightly increase the threshold. Thus, the disease is able to spread more easily in modular networks with communities of similar or equal sizes. With the exception of one data point, this is observed for $\mu > 0.5$ (see the dashed blue line in Fig. 5B).

Real networks. Although the modelling framework presented captures realistic activity and community size distributions of real networks, it neglects other important features such as burstiness[76–80], and more complex temporal/structural correlations[81–85]. It is then crucial grounding the picture emerging from synthetic models with a real world system. To this extent, we consider a temporal and modular network about scientific collaborations in the American Physical Society (APS). We study 96940 scholars connected by 692667 links (see the Supplementary Information for more details)[86]. We focus on ten years of data (January 1997–December 2006) coarse-grained at a time resolution of one month. To single out the effects introduced by communities on contagion processes, we consider also a randomised version of the dataset. Here, the interactions at each time are shuffled, destroying the community structure, but the sequence of activation times for each node and the degree distribution at each time step are preserved[87]. In order to make sure that the randomisation process removes topological structures, we integrate the two networks over all time steps and we use OSLOM[88] to find the communities. The modularity[89] of the real APS network is $Q = 0.6685$, and of its randomised counterpart $Q = 0.0937$. As expected, the degree preserving randomisation reduces the modularity significantly. Using these two networks, we study the dynamical properties of SIR and SIS processes unfolding on their structure. In Fig. 6A,B we present the results. The modular properties of the real network do not influence the threshold of SIR models. Considering the weak dependence on the modularity observed in synthetic networks this results is not surprising. Even more, in this case the maximum value of modularity is defined by the data. We cannot increase it manually as done in our model. Nevertheless, the presence of communities reduces the impact of the disease, i.e. lowers the epidemic size. In the case of SIS processes instead, communities have a larger effect shifting the threshold to smaller values. These results qualitatively confirm what observed in synthetic systems.

Discussion

Real networks are characterised by heterogeneous statistical distributions of crucial topological features; they are organised in modules/communities; they are subject to non trivial temporal dynamics[3–5,8,22,23]. It has long been acknowledged that such attributes have critical effects on contagion processes evolving on systems' fabric[5]. In particular, the heterogeneity in the connectivity patterns makes static/annealed networks extremely fragile to the spreading of infectious diseases[6]. Moreover, the presence of communities in static/annealed graphs might either slow down or facilitate the propagation of a disease[12–18]. In particular, community structures are most likely to inhibit SIR-like processes[12–17]. However, in particular networks, such as Autonomous System Graphs, peculiar topological properties might have the opposite effect[17]. The study of SIS processes in the context of modular networks has received much less attention. However, at least in the case of one artificial network model, modularity has been found to help the spreading[18]. A note of caution is however important. Indeed, in this paper, modular networks are compared with random graphs of different degree distribution. Thus, it is hard to disentangle the real effect of modularity on SIS processes in this case. The temporal features of networks have been found to either facilitate or hamper the spreading of contagion processes. We refer the reader to ref.[90] for a recent compendium of epidemic spreading on time-varying networks. The study of the effects of modularity on epidemic spreading

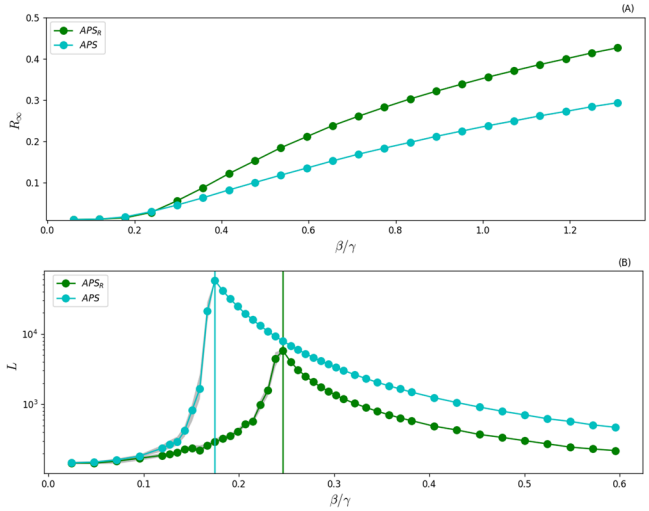


Figure 6. Panel (A) R_∞ as a function of β/γ for SIR processes diffusing on APS (cyan circles) and on the randomized APS dataset (green circles). Panel (B) L as a function of β/γ for a SIS models evolving on the same two networks. Each point is the average of 10^2 independent simulations started from 1% of random seeds. We fix $\gamma = 0.05$.

unfolding on time-varying networks has been very limited. At the best of our knowledge only one paper so far tackled directly this issue[55]. In this work, an activity-driven network organised in two communities of equal size has been considered. In these settings, the threshold is not function of the modularity as each community is a good representation of the full network, thus the critical behaviour is the same independently of the number of connections between the two modules. Furthermore, the authors found a weak correlation between modularity and epidemic size. Starting from all these results, here we aimed to characterise the interplay between modularity and temporal dynamics of networks considering realistic community structures. To this end, we proposed a model of temporal networks with tunable modularity and heterogeneous activity distributions. We provided an analytical description of time-aggregated properties of such networks, and studied the impact of modular and temporal features on epidemic spreading processes. In synthetic networks, we found that modularity reduces the epidemic size and threshold in SIR models, slowing down the spreading process. The effect of modularity on the threshold is weak and appreciable only for large values of it. In SIS models, modularity reduces the epidemic threshold making the system more prone to disease spreading. The dependence on modularity is in this case much stronger. The repetition of the same contacts between nodes belonging to the same community acts as a reservoir for SIS-like diseases and allows the pathogen to reach an endemic state more easily. Our work is not exempt by caveats. The most important limitation, perhaps, is represented by the adoption of Poissonian activation dynamics, which has been shown to be unrealistic in empirical temporal networks which are characterised by bursty behaviours (heterogenous inter-event time distributions)[23]. Furthermore, not only links are dynamical entities being created and terminated, but also nodes may appear and disappear during the dynamics[23]. Our model of modular activity-driven networks does not capture all these aspects of real temporal networks. Thus, we confirmed the picture emerging in synthetic graphs by numerical simulations of SIR and SIS models on real network, characterised by both modular and temporal features. We selected a co-authorship network in which nodes describe authors and links between them capture scientific collaborations. Clearly, such network is not of direct epidemiological relevance. Unfortunately, datasets more suitable for the study of the behaviour of spreading processes on temporal networks, such as face-to-face interactions, are scarce and quite small in size. Thus, they are not the optimal choice to study how temporal and topological properties affect epidemic thresholds which are analytically defined in the limit of $N \to \infty$. However, it is important to notice that social networks, in many different contexts, are characterised by the wide range of features missing in our simple model[3,23]. Thus, in the spirit of a qualitative comparison, studying the behaviour of contagion processes on co-authorship networks is a meaningful exercise. In conclusion, our findings show that on time-varying networks modularity can have opposite effects on different classes of spreading processes. Dynamical processes unfolding on real networks have been show to exhibit a rich and complex phenomenology, depending on the topological and temporal properties of the underlying substrate as well as on the characteristics of the process under investigation. Such complexity may be addressed by both the definition of proper generative models, that allows to control the desired features, and the study of the processes behaviour in real-world network. Our work is set within this framework, and it contributes to shed light on the impact of modular and temporal properties of real networks on epidemic spreading dynamics.

Methods

SIS and SIR models. In both processes nodes are divided in different classes according to their disease status. In SIR models nodes are either Susceptible (S), Infected (I) or Recovered (R). Susceptible nodes describe healthy individuals. Infected nodes contract the disease and are infectious. Recovered nodes are no longer

infected and acquire complete immunity to the illness. The model is fully characterized by two transitions: $S + I \xrightarrow{\beta} 2I$ and $I \xrightarrow{\gamma} R$. The first describes the infection propagation and β is the capita infection rate. This quantity is defined by the average contacts per node $\langle k \rangle$ and by the per contact probability of transmission λ, i.e. $\beta = \lambda\langle k \rangle$. The second transition describes the recovery process. Infected individuals recover spontaneously and permanently with rate γ. In SIS models instead we have just Susceptible and Infected nodes. While the contagion process is equivalent to the SIR case, the recovery is different and described by the following transition: $I \xrightarrow{\gamma} S$. Infected nodes spontaneously return in the susceptible compartment with rate γ.

References

1. Barabási, A.-L. The network takeover. *Nat. Phys.* **8** (2012).
2. Butts, C. Revisiting the foundations of network analysis. *Science* **325**, 414–416 (2009).
3. Newman, M. *Networks. An Introduction* (Oxford University Press, 2010).
4. Caldarelli, G. *Scale-Free Networks* (Oxford University Press, 2007).
5. Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical processes on complex networks* (Cambridge, 2008).
6. Vespignani, A. Modeling dynamical processes in complex socio-technical systems. *Nat. Phys.* **8**, 32–30 (2012).
7. Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function*. (Cambridge University Press, Cambridge, 2010).
8. Fortunato, S. Community detection in graphs. *Phys. Reports* **486**, 75–174 (2010).
9. Colizza, V. & Vespignani, A. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.* **99**, 148701 (2007).
10. Apolloni, A., Poletto, C., Ramasco, J., Jensen, P. & Colizza, V. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theor. biology & medical modelling* **11**, 3 (2014).
11. Buscarino, A., Fortuna, L., Frasca, M. & Rizzo, A. Local and global epidemic outbreaks in populations moving in inhomogeneous environments. *Phys. Rev. E* **90**, 042813 (2014).
12. Onnela, J.-P. *et al*. Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci.* **104**, 7332–7336 (2007).
13. Karsai, M. *et al*. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 025102 (2011).
14. Salathé, M. & Jones, J. H. Dynamics and control of diseases in networks with community structure. *PLoS computational biology* **6**, e1000736 (2010).
15. Huang, W. & Li, C. Epidemic spreading in scale-free networks with community structure. *J. Stat. Mech. Theory Exp.* **2007**, P01014 (2007).
16. Wu, X. & Liu, Z. How community structure influences epidemic spread in social networks. *Phys. A: Stat. Mech. its Appl.* **387**, 623–630 (2008).
17. Stegehuis, C., van der Hofstad, R. & van Leeuwaarden, J. S. Epidemic spreading on complex networks with community structures. *Sci. reports* **6** (2016).
18. Liu, Z. & Hu, B. Epidemic spreading in community networks. *EPL (Europhysics Lett.* **72**, 315 (2005).
19. Centola, D. The spread of behavior in an online social network experiment. *Sci.* **329**, 1194–1197 (2010).
20. Centola, D. & Baronchelli, A. The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proc. Natl. Acad. Sci.* **112**, 1989–1994 (2015).
21. Kawadia, V. & Sreenivasan, S. Sequential detection of temporal communities by estrangement confinement. *Sci. reports* **2**, 794 (2012).
22. Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97 (2012).
23. Holme, P. Modern temporal network theory: a colloquium. *The Eur. Phys. J. B* **88**, 234 (2015).
24. Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of dynamic networks. *Sci. Reports* **2**, 469 (2012).
25. Alessandretti, L., Sun, K., Baronchelli, A. & Perra, N. Random walks on activity-driven networks with attractiveness. *Phys. Rev. E* **95**, 052318 (2017).
26. Barabasi, A.-L. The origin of bursts and heavy tails in human dynamics. *Nat.* **435**, 207–211 (2005).
27. Barrat, A. & Cattuto, C. Face-to-face interactions. In *Social Phenomena*, 37–57 (Springer International Publishing, 2015).
28. Sekara, V., Stopczynski, A. & Lehmann, S. Fundamental structures of dynamic social networks. *Proc. national academy sciences* **113**, 9977–9982 (2016).
29. Ódor, G. Slow, bursty dynamics as a consequence of quenched network topologies. *Phys. Rev. E* **89**, 042102 (2014).
30. Porfiri, M., Stilwell, D. J., Bollt, E. M. & Skufca, J. D. Random talk: random walk and synchronizability in a moving neighborhood network. *Phys. D: Nonlinear Phenom.* **224**, 102–113 (2006).
31. Frasca, M., Buscarino, A., Rizzo, A., Fortuna, L. & Boccaletti, S. Synchronization of moving chaotic agents. *Phys. Rev. Lett.* **100**, 044102 (2008).
32. Frasca, M., Buscarino, A., Rizzo, A. & Fortuna, L. Spatial pinning control. *Phys. Rev. Lett.* **108**, 204102 (2012).
33. Ferreira, S. C., Castellano, C. & Pastor-Satorras, R. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Phys. Rev. E* **86**, 044125 (2012).
34. Frasca, M., Buscarino, A., Rizzo, A., Fortuna, L. & Boccaletti, S. Dynamical network model of infective mobile agents. *Phys. Rev. E* **74**, 036110 (2006).
35. Rocha, L. E. C., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol* **7**, e1001109 (2011).
36. Isella, L. *et al*. What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol* **271**, 166 (2011).
37. Miritello, G., Moro, E. & Lara, R. Dynamical strength of social ties in information spreading. *Phys. Rev. E* **83**, 045102 (2011).
38. Karsai, M., Perra, N. & Vespignani, A. Time varying networks and the weakness of strong ties. *Sci. Reports* **4** (2014).
39. Scholtes, I. *et al*. Slow-down vs. speed-up of information diffusion in non-markovian temporal networks. *arXiv:1307.4030* (2013).
40. Lambiotte, R., Salnikov, V. & Rosvall, M. Effect of memory on the dynamics of random walks on networks. *J. Complex Networks* **3**, 177–188 (2014).
41. Rizzo, A., Frasca, M. & Porfiri, M. Effect of individual behavior on epidemic spreading in activity driven networks. *Phys. Rev. E* **90**, 042801 (2014).
42. Sun, K., Baronchelli, A. & Perra, N. Contrasting effects of strong ties on sir and sis processes in temporal networks. *The Eur. Phys. J. B* **88**, 326 (2015).
43. Rizzo, A. & Porfiri, M. Innovation diffusion on time-varying activity driven networks. *EPJ B* **89**, 1–8 (2016).
44. Rizzo, A., Pedalino, B. & Porfiri, M. A network model for ebola spreading. *J. Theor. Biol.* **394**, 212–222 (2016).
45. Zino, L., Rizzo, A. & Porfiri, M. Continuous-time discrete-distribution theory for activity-driven networks. *Phys. review letters* **117**, 228302 (2016).
46. Speidel, L., Klemm, K., Eguiluz, V. M. & Masuda, N. Temporal interactions facilitate endemicity in the susceptible-infected-susceptible epidemic model. *New J. Phys.* **18**, 073013 (2016).
47. Liu, M.-X. *et al*. Social contagions on time-varying community networks. *Phys. Rev. E* **95**, 052306 (2017).
48. Artime, O., Ramasco, J. J. & San Miguel, M. Dynamics on networks: competition of temporal and topological correlations. *Sci. Reports* **7** (2017).

49. Stehle, J. *et al*. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One* **6**, e23176 (2011).
50. Liu, S., Perra, M., Karsai, N. & Vespignani, A. Controlling contagion processes in activity driven networks. *Phys. Rev. Lett.* **112**, 118702 (2014).
51. Keeling, M. & Rohani, P. *Modeling Infectious Disease in Humans and Animals* (Princeton University Press, 2008).
52. Tomasello, M. V., Perra, N., Tessone, C. J., Karsai, M. & Schweitzer, F. The role of endogenous and exogenous mechanisms in the formation of r&d networks. *Sci. reports* **4** (2014).
53. Ribeiro, B., Perra, N. & Baronchelli, A. Quantifying the effect of temporal resolution on time-varying networks. *Sci. Reports* **3**, 3006 (2013).
54. Lancichinetti, A., Fortunato, S. & Radicchi, F. New benchmark in community detection. *Phys. Rev. E* **78**, 046110 (2008).
55. Han, D., Sun, M. & Li, D. Epidemic process on activity-driven modular networks. *Phys. A: Stat. Mech. its Appl.* **432**, 354–362 (2015).
56. Starnini, M. & Pastor-Satorras, R. Topological properties of a time-integrated activity-driven network. *Phys. Rev. E* **87**, 062807 (2013).
57. Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
58. Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J. G. & Mendes, J. F. F. Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.* **109**, 128702 (2012).
59. Sun, K., Baronchelli, A. & Perra, N. Epidemic spreading in non-markovian time-varying networks. *arxiv:1404.1006* (2014).
60. Wang, Y., Chakrabarti, D., Wang, G. & Faloutsos, C. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proc 22nd Int. Symp. on Reliab. Distributed Syst.* 25–34 (2003).
61. Durrett, R. Some features of the spread of epidemics and information on a random graph. *Proc. Nat. Acad. Sci.* **107**, 4491–4498 (2010).
62. Prakash, B., Tong, H., Valler, M. & Faloutsos, C. Virus propagation on time-varying networks: Theory and immunization algorithms. *Mach. Learn. Knowl. Discov. Databases Lect. Notes Comput. Sci.* **6323**, 99–114 (2010).
63. Valdano, E., Ferreri, L., Poletto, C. & Colizza, V. Analytical computation of the epidemic threshold on temporal networks. *Phys. Rev. X* **5**, 021005 (2015).
64. Starnini, M., Machens, A., Cattuto, C., Barrat, A. & Pastor-Satorras, R. Immunization strategies for epidemic processes in time-varying contact networks. *J. Theor. Biol.* **337**, 89–100 (2013).
65. Lee, S., Rocha, L., Liljeros, F. & Holme, P. Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS One* **7**, e36439 (2012).
66. Takaguchi, T., Sato, N., Yano, K. & Masuda, N. Importance of individual events in temporal networkss. *New J. Phys.* **14**, 093003 (2012).
67. Tang, J., Mascolo, C., Musolesi, M. & Latora, V. Exploiting temporal complex network metrics in mobile malware containment. In *Proceedings of IEEE 12th International Symposium on a World of Wireless, Mobile and Multimedia Networks* (2011).
68. Masuda, N. & Holme, P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Reports* **5** (2013).
69. Pozzana, I., Sun, K. & Perra, N. Epidemic spreading on activity-driven networks with attractiveness. *arXiv preprint arXiv:1703.02482* (2017).
70. Liu, S., Baronchelli, A. & Perra, N. Contagion dynamics in time-varying metapopulations networks. *Phy. Rev. E* **87** (2013).
71. Starnini, M. & Pastor-Satorras, R. Temporal percolation in activity driven networks. *Phys. Rev. E* **89**, 032807 (2014).
72. Morris, M. *Sexually Transmitted Diseases*, (Holmes, K. K. *et al*. Eds) (McGraw-Hill, 2007).
73. Onaga, T., Gleeson, J. & Masuda, N. Concurrency-induced transitions in epidemic dynamics on temporal networks. *Phys. Rev. Lett.* **108301** (2017).
74. Boguña, M., Castellano, C. & Pastor-Satorras, R. Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks. *Phys. Rev. Lett.* **111**, 068701 (2013).
75. Aharony, A. & Stauffer, D. *Introduction to percolation theory* (Taylor & Francis, 2003).
76. Ubaldi, E., Vezzani, A., Karsai, M., Perra, N. & Burioni, R. Burstiness and tie activation strategies in time-varying social networks. *Sci. Reports* **7** (2017).
77. Goh, K.-I. & Barabási, A.-L. Burstiness and memory in complex systems. *EPL (Europhysics Lett.* **81**, 48002 (2008).
78. Moinet, A., Starnini, M. & Pastor-Satorras, R. Burstiness and aging in social temporal networks. *Phys. review letters* **114**, 108701 (2015).
79. Lambiotte, R., Tabourier, L. & Delvenne, J.-C. Burstiness and spreading on temporal networks. *The Eur. Phys. J. B* **86**, 320 (2013).
80. Karsai, M., Kaski, K., Barabási, A.-L. & Kertész, J. Universal features of correlated bursty behaviour. *Sci. reports* **2** (2012).
81. Peixoto, T. & Rosvall, M. Modelling sequences and temporal networks with dynamic community structures. *Nat. Commun.* **8** (2017).
82. Laurent, G., Saramäki, J. & Karsai, M. From calls to communities: a model for time-varying social networks. *The Eur. Phys. J. B* **88**, 301 (2015).
83. Pfitzner, R., Scholtes, I., Garas, A., Tessone, C. J. & Schweitzer, F. Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *Phys. review letters* **110**, 198701 (2013).
84. Vestergaard, C. L., Génois, M. & Barrat, A. How memory generates heterogeneous dynamics in temporal networks. *Phys. Rev. E* **90**, 042805 (2014).
85. Ubaldi, E. *et al*. Asymptotic theory of time-varying social networks with heterogeneous activity and tie allocation. *Sci. reports* **6** (2016).
86. Radicchi, F., Fortunato, S., Markines, B. & Vespignani, A. Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E* **80**, 056103 (2009).
87. Starnini, M., Baronchelli, A., Barrat, A. & Pastor-Satorras, R. Random walks on temporal networks. *Phys. Rev. E* **85**, 056115 (2012).
88. Lancichinetti, A., Radicchi, F., Ramasco, J. J. & Fortunato, S. Finding Statistically Significant Communities in Networks. *Plos One* (2011).
89. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69** (2004).
90. Masuda, N. & Holme, P. Temporal network epidemiology (2017).

Acknowledgements

M.S. acknowledges financial support from the James S. McDonnell Foundation. M.N. thanks the Centre for Business Networks Analysis at the University of Greenwich for support and hospitality during this project. M.N. and A.R. acknowledges financial support from the National Science Foundation under grant No. CMMI-1561134 and the Army Research Office under grant No. W911NF-15-1-0267, with Drs A. Garcia and S.C. Stanton as program managers. A.R. acknowledges financial support from Compagnia di San Paolo, Italy.

Author Contributions

N.P. conceived the research, M.N. and K.S. conducted the numerical simulations, E.U. developed the analytical calculations, all authors analysed the results, wrote and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-20908-x.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information: epidemic spreading in modular time-varying networks

Matthieu Nadini[1,6], Kaiyuan Sun[2], Enrico Ubaldi[3], Michele Starnini[4,5], Alessandro Rizzo[6], and Nicola Perra[7,*]

[1]Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn NY 11201, USA
[2]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston MA 02115 USA
[3]Institute for Scientific Interchange, ISI Foundation, Turin, Italy
[4]Departament de Física Fondamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain
[5]Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain
[6]Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[7]Centre for Business Networks Analysis, University of Greenwich, London, UK
[*]n.perra@greenwich.ac.uk

ABSTRACT

In this Supplementary Information, we present in a more detailed and comprehensive way the definition of the model, its key properties and the results (both analytical and numerical) found. Furthermore, we show additional simulations studied in synthetic networks for SIR and SIS models. Finally, an explanation of the APS dataset is given.

1 The Model

The network is defined by means of the following parameters:

- the total number N of nodes in the network;
- the activity distribution parameters, i.e. the lower cut-off ε and the leading exponent ν, so that $F(a) \propto a^{-\nu}$ for $a \in [\varepsilon, 1]$;
- the lower cut-off $s_{\min}$, the upper limit $s_{\max}$ and the exponent ω describing the community size distribution, i.e. $P(s) \propto s^{-\omega}$ for $s \in [s_{\min}, s_{\max}]$;
- μ is the probability that, once active, a node will connect to a node inside the community, so that $\mu' = 1-\mu$ is the probability to fire outside the community;

We initialize the network extracting N activity values from the activity distribution $F(a)$ and we then group the nodes in communities of size s drawn from a size-distribution $P(s)$. Once we initialized the network we let it evolve following the time-varying activity driven framework. At each time step t we start with N disconnected nodes. Each node gets active with probability $a_i dt$ at each time step dt and fire to a randomly chosen node inside (outside) its own community with probability μ (μ'). At time $t+dt$ we delete all the edges and repeat the above procedure. Each node i will then have a set of neighbors that have been contacted or have contacted the node during the network growth. The size of such a set is the integrated degree k of the node i. Of these k neighbors, k_c will be inside the i's community (in-community degree) and $k_o = k - k_c$ will be external to the community (out-community degree).

2 The Network growth

In the integrated network, each node i has a set of neighbors that have been contacted or have contacted the node during the network growth. The size of such a set is the integrated degree k. Of these k neighbors, k_c are inside the i's community (i.e. the in-community degree) and $k_o = k - k_c$ are external to the community (i.e. the out-community degree). Since the model is memoryless, the in-community degree k_c and the out-community degree k_o are decoupled and can, in fact, be treated separately. Even the activity potential a_i of each node can be "split" in two components: the in-community activity $\mu a_i = (1-\mu')a_i$

and the complementary out-community $\mu' a_i$. Indeed, each node points, on average, a fraction μ of its own events toward the community, while the remaining μ' are directed outside the community itself. Each node experiences a mean field of activity, $\mu \langle a \rangle$, coming from the community (provided that the community is large enough) and a supplementary external field $\mu' \langle a \rangle$ coming from the rest of the network.

2.1 The network time scales

As a first insight, let us note that the in-degree time dependence can be easily approximated with a probabilistic consideration. Each node i of activity a_i, within a community of size s, has $s-1$ available edges. Now, for each time step of the dynamics, the edge e_{ij} is created with probability $\mu(a_i + a_j)/(s-1)$. Then, on average, each edge emanating from i is activated with probability $c(a_i,\mu)/(s-1) = \mu(a_i + \langle a \rangle)/(s-1)$, where $c(a_i,\mu)$ is the number of edges intra-community. The probability $P(a_i,\mu,s,t)$ for an edge pointing to i not to be activated after t time steps then reads:

$$P(a_i,\mu,s,t) = \left(1 - \frac{c(a_i,\mu)}{s-1}\right)^t. \tag{1}$$

Since the in-degree $k_c \in [0, s-1]$, we can write:

$$k_c(a_i,\mu,s,t) = (s-1)(1 - P(a_i,\mu,s,t)). \tag{2}$$

In the following, we always have the dependency on a_i, μ and t, so to simplify the notation we drop most of those parameters: $P(a_i,\mu,s,t) = P'(s)$, to avoid confusions with the community size distribution P(s). Also, $k_c(a_i,\mu,s) = k_c(s)$ and $c(a_i,\mu) = c$.

We note that Eq. 1 gives us an estimation of the characteristic time $\tau(s)$ that takes for a node of activity a_i to saturate the in-degree $k_c \to (s-1)$. Indeed, we can rewrite Eq. 1 as:

$$P'(s) = \exp\left[t \ln\left(1 - \frac{c}{s-1}\right)\right] \Rightarrow \tau(s) = -\left[\ln\left(1 - \frac{c}{s-1}\right)\right]^{-1}. \tag{3}$$

So, as expected, the saturation time (i.e. the typical time for k_c to be of the same order of s) increases as the activity μa_i decreases and/or the community size s grows.

Generalizing the above reasoning, the characteristic time for a community to have the majority of the nodes saturated is obtained by evaluating the probability $P_e(s)$ to create (on average) an edge e_{ij} in a community of size s in a single evolution step. In other words, it is the number of edges activated in one step divided by the total number of possible edges in the network:

$$P_e(s) = \frac{2s\mu \langle a \rangle}{s(s-1)}. \tag{4}$$

The probability for one edge not to be created after t time steps is then:

$$\bar{P}_e(s) = \left(1 - \frac{2\mu \langle a \rangle}{s-1}\right)^t \Rightarrow \bar{P}_e(s) = \exp\left[-\frac{t}{\tau_c(s)}\right] \Rightarrow \tau_c(s) = -\left[\ln\left(1 - \frac{2\mu \langle a \rangle}{s-1}\right)\right]^{-1}, \tag{5}$$

where $\tau_c(s)$ represents the typical time by which the majority of the nodes of a community has a degree $k_c \simeq s$.

Note that, in the evaluation of both $\tau(s)$ and $\tau_c(s)$, we did not take into account the difference between edges pointing to a more active node and the ones pointing to a less active one. Nevertheless, this is a simple estimation that, as we will show later, correctly catches the general behaviour of the in-degree k_c for any value of μ, a_i and s. Besides, when computing the key features of the evolving network, we are now able to distinguish the short time range $t \ll \tau_c(s)$ (in which $k_c \ll s$ for any activity value a_i) and the long time limit $t \gg \tau_c(s)$ (in which $k_c \sim s$ for any activity value a_i).

2.2 The Master Equation and the $P(a,k,t)$

We can now write down the Master Equation (ME) for the quantities $P_c(s,k_c)$ and $P_o(s,k_o)$, that is, the probability for a node of activity a_i belonging to a community of size s to have degree in (out) degree k_c (k_o) at time t. In general, $a_i \Delta t$ represents the probability the node i is active, where a_i is the activity rate of node i. Without loss of generality we will assume $\Delta t = 1$. To get the ME for the in-degree k_c distribution, we exploit also the time-dependence for a couple of passages:

$$\begin{aligned} P_c(s,k_c,t+1) = {} & P_c(s,k_c,t)\left[1 - \mu \sum_j a_j\right] + P_c(k_c,t)\left[\frac{k_c}{s}\mu a_i + \frac{s-1}{s}\mu \sum_{j \nsim i} a_j + \mu \sum_{j \sim i} a_j\right] + \\ & + P_c(s,k_c-1,t)\left[\frac{s-k_c}{s}\mu a_i + \frac{1}{s}\mu \sum_{j \nsim i} a_j\right], \end{aligned} \tag{6}$$

where $\sum_{j\sim i}$ and $\sum_{j\nsim i}$ are respectively two contracted notations for the sum of all the first neighbors of node i and the sum of all nodes but the neighbors of node i. The first parenthesis indicates the probability that none of the nodes in the network fire. Third (sixth) term is the probability a node i, in the instantaneous network, is active and fires to a node where, in the integrated counterpart, there is already (isn't) a link. Four (seventh) term is the probability a node j not linked to i fires to any of the nodes but i (fires to i). Fifth factor is the probability a node j already linked to i fires.
After some algebra, ME can be written as:

$$P_c(s,k_c,t+1)-P_c(s,k_c,t)=-\left[P_c(s,k_c,t)-P_c(s,k_c-1,t)\right]\left(\frac{s-k_c}{s}\mu a_i+\frac{\mu}{s}\sum_{j\nsim i}a_j\right). \tag{7}$$

Now we pass to the continuum limit by considering $t\gg 1$ and $k\gg 1$. So the *l.h.s* becomes simply the time derivative with respect to $P_c(s,k_c)$ and, to obtain a proper convergence of the results, we can expand the probability with respect to the incommunity degree up to second order.
In the regime $t\ll\tau(s)$, we can neglect $k_c\ll s$ and $1/s\sum_{j\nsim i}a_j\approx\langle a\rangle$.

$$\frac{\partial P_c(s,k_c)}{\partial t}=(\mu a+\mu\langle a\rangle)\left[\frac{\partial P_c(s,k_c)}{\partial k_c}-\frac{1}{2}\frac{\partial^2 P_c(s,k_c)}{\partial k_c^2}\right], \tag{8}$$

where we dropped the a_i index since we expect all the nodes of a given activity to behave in the same way. Now a is an activation rate and in the treatment we assume it takes small values to avoid that two nodes become active together.
The solution of Eq. 8 reads:

$$P_c(s,k_c)=C\exp\left[-\frac{(k_c-\mu(a+\langle a\rangle)t)^2}{2\mu(a+\langle a\rangle)t}\right], \tag{9}$$

where C is a normalization constant.
By following the same procedure, we recover the same results of Eq. 9 for the out-community degree $k_o(a,t)$, by substituting $\mu\to\mu'$ and $k_c\to k_o$:

$$P_o(s,k_o)=C\exp\left[-\frac{(k_o-\mu'(a+\langle a\rangle)t)^2}{2\mu'(a+\langle a\rangle)t}\right] \tag{10}$$

Since $N\gg s_{max}$, the out-degree $k_o\ll N$ for any time t of the process, thus we assume that Eq. 10 is valid for all the time scales analyzed. Also note that, as expected, the net effect of the mixing parameter μ' is just a time rescaling of the out-community and in-community activity, respectively.
Then, in the $t\sim\tau_c(s)$ time range is not possible to find an analytic formula, however simulations will be run to provide, at least, a qualitative behavior. In the $t\gg\tau_c(s)$ time limit the $P_c(s,k_c)$ converges to the $\delta(k_c-(s-1))$ distribution. In fact, all the nodes will have all their edges activated and the $P_c(s,k_c)$ time derivative goes to zero.

Let us now resume the results found in this section:

$$P_c(s,k_c)\propto\begin{cases}\exp\left[-\dfrac{(k_c-\mu(a+\langle a\rangle)t)^2}{2\mu(a+\langle a\rangle)t}\right] & \text{for } t\ll\tau_c(s) \quad (11a)\\ \delta(k_c-(s-1)) & \text{for } t\gg\tau_c(s) \quad (11b)\end{cases}$$

$$P_o(s,k_o)\propto\exp\left[-\frac{(k_o-\mu'(a+\langle a\rangle)t)^2}{2\mu'(a+\langle a\rangle)t}\right]\quad\text{for }\forall t \tag{12}$$

where we now distinguish between the in-community degree distribution $P_c(s,k_c)$ and the out-community degree distribution $P_o(s,k_o)$. The latter however, is independent on the community size and we can then define $P_o(s,k_o)=P_o(k_o)$.
So far we treated the two probability functions separately when, in fact, k_c and k_o are bound by the relation $k=k_c+k_o$. The total degree distribution $P(s,k)$ will then be determined by the convolution of both the $P_c(s,k_c)$ and $P_o(s,k-k_c)$:

$$P(s,k)=\int_0^k dk_c P_c(s,k_c)P_o(k-k_c) \tag{13}$$

where we integrate over all the possible arrangements of the k_c edges.
In the $t \ll \tau(s)$ limit, by substituting Eq. 11a and 12 in Eq. 13, we sum the two exponents getting:

$$P(s,k) = C \int_0^k dk_c \exp\left[-\frac{(k_c - \mu(a+\langle a\rangle)t)^2}{2\mu(a+\langle a\rangle)t} - \frac{(k-k_c-\mu'(a+\langle a\rangle)t)^2}{2\mu'(a+\langle a\rangle)t} \right] \tag{14}$$

where C is, again, a normalization constant.
By combining the two terms and after some algebra we get:

$$P(s,k) = C \int_0^k dk_c \exp\left[-\frac{(k_c - \mu k)^2}{2\mu\mu'(a+\langle a\rangle)t} - \frac{k^2 - 2k(a+\langle a\rangle)t - (a+\langle a\rangle)^2 t^2}{2(a+\langle a\rangle)t} \right]. \tag{15}$$

The integration over k_c gives:

$$P(s,k) = C \left[\mathrm{Erf}\left(\frac{\mu' k}{\sqrt{2\mu\mu'(a+\langle a\rangle)t}} \right) - \mathrm{Erf}\left(\frac{\mu k}{\sqrt{2\mu\mu'(a+\langle a\rangle)t}} \right) \right] \exp\left[-\frac{k^2 - 2k(a+\langle a\rangle)t - (a+\langle a\rangle)^2 t^2}{2(a+\langle a\rangle)t} \right] \tag{16}$$

where $\mathrm{Erf}(x)$ is the error function evaluated at x.
In the small time limit, if we want to evaluate the $P(k) = \int_{s_{\min}}^{s_{\max}} ds P(s) P(s,k)$, we have to consider $t \ll \min_s(\tau_c(s))$. In this way, for each community we can use Eq. 16 as the true value of the $P(s,k)$. The integration over the different community size s is then straightforward since the terms are independent on it, giving $P(k) = P(s,k)$. Note that this result holds for any value of μ, N and a.
The computation of $P(k)$ in the large time limit (i.e. $t \gg \langle \tau_c(s) \rangle$) is more complicated and we have to assume that $k_c = s-1$ for each node in a community of size s, otherwise Eq. 11b put everything equal to zero. The $P_o(s,k_o)$ will still be approximated by Eq. 12. The integral now reads:

$$P(k) = \int_{s_{\min}}^{s_{\max}} ds P(s) P(s,k) = C \int_{s_{\min}}^{s_{\max}} ds P(s) \exp\left[-\frac{(k-(s-1)-\mu'(a+\langle a\rangle)t)^2}{2\mu'(a+\langle a\rangle)t} \right], \tag{17}$$

The exponential can be written as:

$$\exp\left[-\frac{(k-\mu'(a+\langle a\rangle)t)^2 + (s-1)[(s-1) - 2(k-\mu'(a+\langle a\rangle)t))]}{2\mu'(a+\langle a\rangle)t} \right], \tag{18}$$

where the rise of new terms proportional to s^2 and sk makes it difficult to perform the integral.

We can, however, give a solution for the simple case $P(s) = \delta(s-\bar{s})$, when all the communities have equal size. First of all, we have, for large times, $P_c(s,k_c) = \delta(k_c - (s-1))$. Then:

$$P(s,k) = \begin{cases} P_o(k-(s-1)) & \text{for } k \geq s-1 \quad (19a) \\ 0 & \text{for } k < s-1. \quad (19b) \end{cases}$$

When $k < s-1$, for sure $k_c < s-1$ and the delta put everything equal to zero. In the other case $k_c = s-1$ with a certain probability $P_o(k-(s-1))$. Finally, equation 17 can be written as:

$$P(k) = \begin{cases} P_o(k-(\bar{s}-1)) & \text{for } k \geq \bar{s}-1 \quad (20a) \\ 0 & \text{for } k < \bar{s}-1 \quad (20b) \end{cases}$$

2.3 The average degree $\langle k(a,s,t) \rangle$

We can provide a simple expression for the nodes average degree belonging in different classes. As we already showed in Eq. 2, $\langle k_c(a,s,t) \rangle$ grows as:

$$\begin{aligned} \langle k_c(a,s,t) \rangle &= (s-1)\left(1 - \exp\left(-\frac{t}{\tau(s)}\right)\right) = \\ &= (s-1)\left(1 - P'(s)\right) = \\ &= (s-1)\left[1 - \left(1 - \frac{c}{s-1}\right)^t\right], \end{aligned} \tag{21}$$

and the $\langle k_o \rangle$ grows as the mean value of the distribution given in equation 12, and turns out to be independent on s.

$$\langle k_o(a,t) \rangle = \mu'(a + \langle a \rangle)t \tag{22}$$

The average total degree $k(a,s,t)$ for nodes of activity a belonging to communities of size s depends on the time scale we analyse the problem. For small times (i.e. $t \ll \langle \tau(s) \rangle_s$):

$$\begin{aligned} \langle k(a,s,t) \rangle &= \langle k_c(a,s,t) \rangle + \langle k_o(a,t) \rangle \approx \\ &\approx (s-1)\left(1 - 1 + \frac{\mu(a+\langle a \rangle)t}{s-1}\right) + (1-\mu)(a+\langle a \rangle)t = \\ &= (a + \langle a \rangle)t \end{aligned} \tag{23}$$

As time grows toward the regime of times comparable to the average (i.e. $t \sim \langle \tau(a,s,t) \rangle_s$), we cannot approximate the in-community degree anymore but we use directly equation 2:

$$\langle k(a,s,t) \rangle = (s-1)(1 - P'(s)) + \langle k_o(a,t) \rangle . \tag{24}$$

Then, the regime of large times (i.e. $t \gg \langle \tau(s) \rangle_s$) is:

$$\langle k(a,s,t) \rangle \approx s - 1 + \mu'(a + \langle a \rangle)t \approx \mu'(a + \langle a \rangle)t \tag{25}$$

The above equations then predict that the average degree has a linear growth proportional to $a + \langle a \rangle$ for short time limit (equation 23), then a transition for $t \sim \langle \tau(s) \rangle_s$ is followed by a second linear growth, valid for large times, proportional to $\mu'(a + \langle a \rangle)$. These regimes correspond to the initial growth, in which both the in-community and the out-community degrees are growing linearly in time, followed by the slowing down of the in-community degree which is saturating to $s-1$. Finally, the third regime is again linear and it is driven by the $\mu'(a + \langle a \rangle)$ coefficient: it means meaning that only the out-community degree is growing.

2.4 The degree distribution $\rho(k_{\text{th}})$

Now that we have the expression of the average degree, it is straightforward to write the degree distribution. At all the time scales we found $\langle k_{\text{th}} \rangle \propto Ct$, where C is a time-independent coefficient. Then, $k_{\text{th}} \propto at$, and it results in an equal increment in the activity and in the degree values ($da = dk_{\text{th}}$). If we use the change of variable rule, we obtain:

$$F(a)da = \rho(k_{\text{th}})dk_{\text{th}} \Rightarrow a^{-\nu}da = k_{\text{th}}^{-x}dk_{\text{th}} \Rightarrow x = \nu, \tag{26}$$

i.e., the degree distribution has the same exponent ν as the activity distribution function.

3 Comparison with numerical simulations

To check the analytical predictions of Section 2 we performed numerical simulation. In particular we realized 100 representations of a network featuring:

- $N = 10^5$ nodes with modularity $\mu = 0.9$ evolving for 10^5 evolution steps;
- activity potential distributed following the $F(a) \propto a^{-\nu}$ with $\nu = 2.1$ and $a \in [10^{-3}, 1]$ interval;
- power-law distributed community sizes $P(s) \propto s^{-\omega}$ with $\omega = 2.1$ and $s \in [10, \sqrt{N}]$.

In order to analyze the collective behavior of the nodes we group them by their activity and community size, thus defining b classes of nodes. We average over the representations of the network and for each class of nodes b we evaluate:

- $P_c(s,k)$, $P_o(k)$, $P(s,k)$ for $t \ll \tau_c(s)$ and $t \gg \tau_c(s)$;
- the average degree $\langle k_c(s) \rangle$, $\langle k_o \rangle$ and $\langle k(s) \rangle$;
- the degree distribution $\rho(k_{\text{th}})$.

In the main discussion, we already showed that some of the above measures have a great agreement with analytical predictions. To complete the discussion, we add below Fig. 1 which proves that also the in-community probability and the in-community degree perfectly matches our expectation. In panel B) we display $P_c(s,k)$ for $t \sim \tau_c$ even if it was impossible to obtain an exact result, the comparison demonstrates that the in-community probability starts to deviate from a Gaussian distribution.

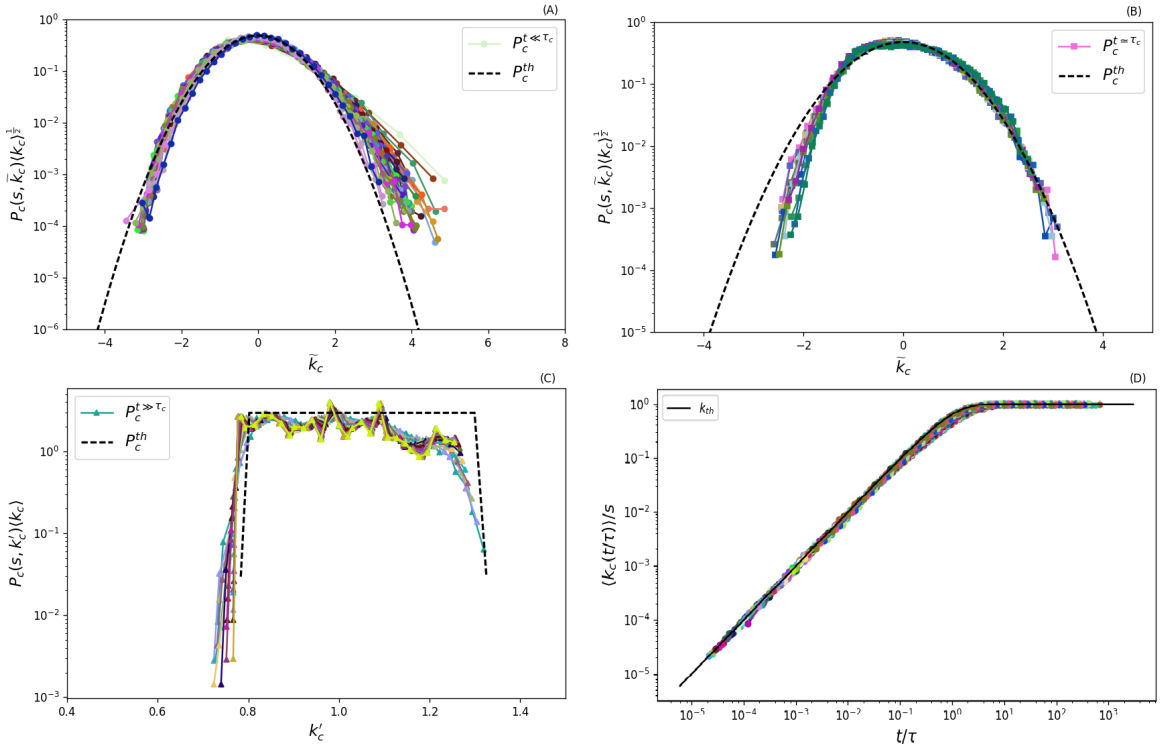


Figure 1. Panels A-C) rescaled $P_c(k_c)$ probability distribution as found for a selected node class at different times (legends). Functions in panels A-B) are rescaled accordingly to the theoretical distribution given in Eq. 11a, i.e., by sending $k_c \to \tilde{k_c} = (k_c - \langle k_c \rangle)/\langle k_c \rangle^{1/2}$ and plotting $P_c(\tilde{k_c})\langle k_c \rangle^{1/2}$. Panel C) is rescaled accordingly to the theoretical distribution given in Eq. 11b, i.e., by sending $k_c \to k'_c = k_c/\langle k_c \rangle$ and plotting $P_c(k'_c)\langle k_c \rangle$. In panel D) we plot $\langle k_c \rangle$ for nodes featuring different activity potential and belonging to communities of different size. The data are rescaled sending the time $t \to = t/\tau$ and then plotting $\langle k_c(t/\tau)\rangle/s$. The analytical prediction of Eq. 2 computed for a node of activity $a = \langle a \rangle$ and belonging to a community of size $s = \langle s \rangle$. Each point is an average over 10^2 independent simulations, parameters used are: $N = 10^5$, $\omega = 2.1$, $\nu = 2.1$, $m = 1$, $s_{min} = 10$ and $\mu = 0.9$

4 SIR and SIS processes on modular activity driven networks

We present together all the results about SIR and SIS models obtained in synthetic networks. We start our simulations by setting 1% of randomly selected nodes as initial infected seeds, the other parameters are fixed as $\gamma = 0.01$, $m = 3$, $\nu = 2.1$, $N = 10^5$ and $Y = 0.5$. Panels A) and C) set the exponent of the distribution of community sizes $\omega = 1.5$, while panels B) and D) have $\omega = 2.1$. The qualitative picture is unchanged due to selecting a different value of ω. The modular structure becomes irrelevant for $\mu \to 0$, whilst it significantly modifies the spread of the disease when $\mu \gg 0$. Modularity slows down contagion processes in SIR models, it favors the disease outbreak in SIS models. Moreover, quantitatively, in SIR models the presence of larger communities lead to higher epidemic size, in SIS models it lower the epidemic threshold. Moreover, in SIS models and when $\mu \to 1$, the epidemic threshold is likely to increase due to major limitations in reaching an endemic size. This phenomenon is particularly visible in Panel C), solid blue curve.

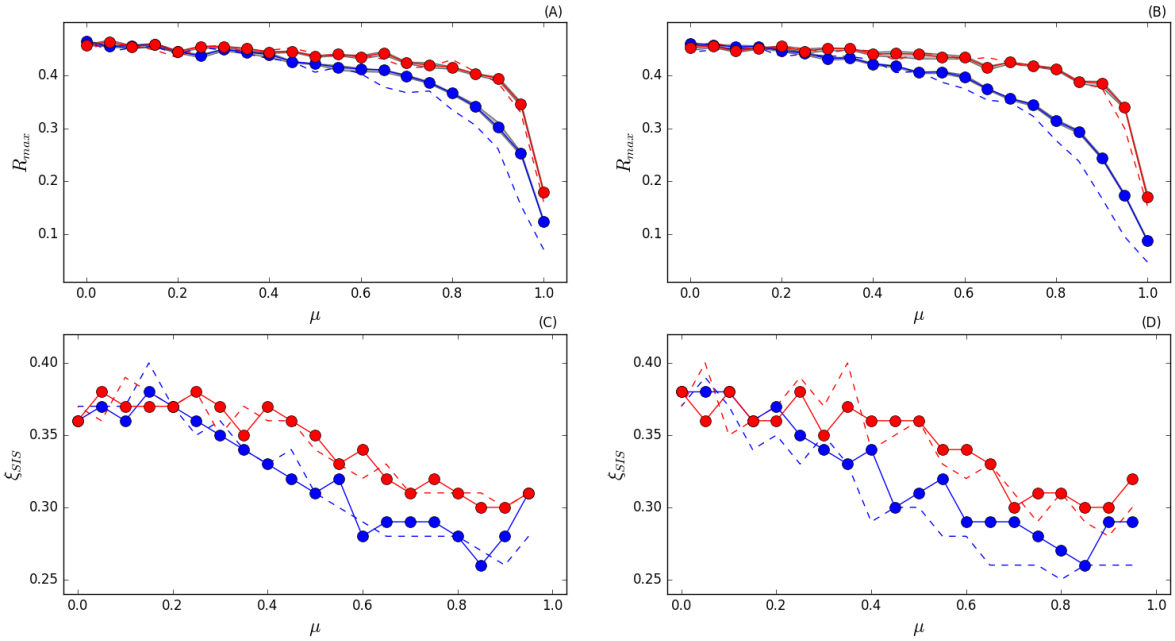


Figure 2. Panels A-B) take R_{max} of each R_∞ curve and plot it as a function of μ. Panels C-D) ξ_{SIS}, that is β/γ in correspondence of L_{max}, and plot it as a function of the modularity. In red curves we set $s_{min} = 100$, in blue curves $s_{min} = 10$. Solid curves are obtained by drawing community sizes from a power law distribution, 95% confidence interval is in gray. Dashed curves have a constant community size equal to the average value of the power law. Each point is an average of 10^2 independent simulations.

5 Real network: APS dataset

In the data each author of an article is described as a node. An undirected link between two different authors is drawn if they collaborated in the same article. We used the dataset from Ref[1] which spans a period between 1893 and 2006. To have the average degree in each instantaneous network as comparable as possible (see Fig. 3), we select a period of ten years, from January 1997 to December 2006. In this time window we register 96940 scholars who create 692667 connections. When we simulate SIR and SIS models on top of APS temporal network, we use periodic boundary conditions to let the disease dynamics evolve without late-time constrains.

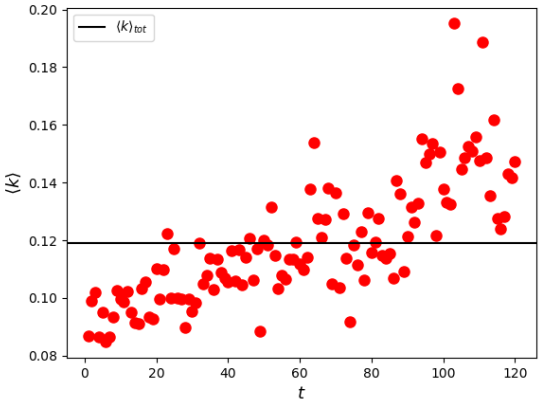


Figure 3. Average degree for each month in the selected subset of APS dataset (red circles), compared with the global average of all the ten years considered (solid black line). We label each month with increasing integer numbers from 1 to 120, where 1 represents the beginning of our sample, January 1997, and 120 the end, December 2006. We observe an increasing number of collaborations through years.

Using OSLOM, in Fig.4A we show that the integrated APS network is modular. Then, we apply the following degree-preserving randomization technique to destroy the network's community structure. We choose randomly a source node S_1 and, among its neighbors, we select randomly a target node T_1. We do the same for other two nodes S_2 and T_2. If the two pairs are equivalent or if a multi-edge will be created, we start back by selecting S_1 again and repeating the instructions. Otherwise, we swap T_1 and T_2 to have the new undirected links: $S_1 - T_2$ and $S_2 - T_1$. The edges are chosen within the same temporal network and the number of swaps is equal to the number of edges in that instantaneous network. So, the above procedure is applied for each instantaneous network. At the end, we integrate the randomized temporal network and use OSLOM to detect the community structure. In Fig.4B, we qualitatively prove that our degree preserving randomization destroys the network's community structure.
The options used in OSLOM are: -uw (to study undirected networks); -cp0.99 (to have communities as large as possible); -hr (to avoid to consider hierarchies); -r100 (to repeat 100 times the community detection). Since OSLOM finds communities in a non-deterministic way, last option is useful to get rid of stochastic fluctuations and have a more reliable community structure. Finally, we evaluate quantitatively the modularity of the two networks. For the original APS network, $Q = 0.6685$, and for its randomized counterpart $Q = 0.0937$.

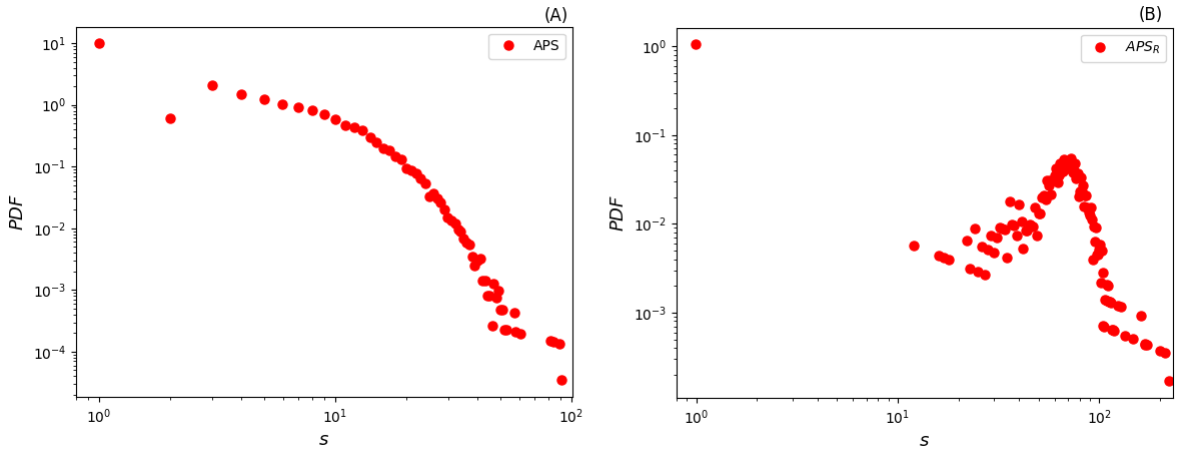


Figure 4. Panel A) Community sizes probability density function in the original network. Number of communities found is 10825, minimum community size is 1, maximum 91. Panel B) Community sizes probability density function after having applied the degree-preserving randomization. Number of communities found is 1489, minimum community size is 1, maximum 222. Note also that the shape of the distribution is completely different from panel A).

References

1. Radicchi, F., Fortunato, S., Markines, B. & Vespignani, A. Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E* **80**, 056103 (2009).