Diffusion on networked systems is a question of time or structure

Jean-Charles Delvenne*

ICTEAM and CORE, University of Louvain, 4 Avenue Lemaître, B-1348 Louvain-la-Neuve, Belgium

Renaud Lambiotte

Department of Mathematics and naXys, Université de Namur, 8 Rempart de la Vierge, B-5000 Namur, Belgium

Luis E C Rocha

Department of Mathematics and naXys, Université de Namur, 8 Rempart de la Vierge, B-5000 Namur, Belgium
Department of Public Health Sciences, Karolinska Institutet, 18A Tomtebodavägen, S-17177 Stockholm, Sweden

Network science investigates the architecture of complex systems to understand their functional and dynamical properties. Structural patterns such as communities shape diffusive processes on networks. However, these results hold under the strong assumption that networks are static entities where temporal aspects can be neglected. Here we propose a generalised formalism for linear dynamics on complex networks, able to incorporate statistical properties of the timings at which events occur. We show that the diffusion dynamics is affected by the network community structure and by the temporal properties of waiting times between events. We identify the main mechanism — network structure, burstiness or fat-tails of waiting times — determining the relaxation times of stochastic processes on temporal networks, in the absence of temporal-structure correlations. We identify situations when fine-scale structure can be discarded from the description of the dynamics or, conversely, when a fully detailed model is required due to temporal heterogeneities.

The relation between network structure and dynamics has attracted the attention of researchers from different disciplines over the years[1–5]. These works are rooted in the observation that, in contexts as diverse as the Internet, society, and biology, networks tend to possess complex patterns of connectivity, with a significant level of heterogeneity[4]. In addition, a broad range of real-world dynamical processes, from information to virus spreading, is akin to diffusion. If the effect of structure, such as communities or degree heterogeneity, on diffusive processes is now well-known[6,7], the impact of the temporal properties of individual nodes is poorly understood[1,8]. Yet, empirical evidence indicates that real-world networked systems are often characterised by complex temporal patterns of activity, including a fat-tailed distribution of times[9–16], correlations between events[17,18] and non-stationarity[19–21]. This is in remarkable contrast to a vast majority of mathematical models, which assume homogeneous interaction dynamics on networks[1,2,22,23]. In this work, we focus on the impact of structure and the waiting time distribution on dynamics, and set aside any other temporal properties. This subject has triggered intense theoretical work in recent years, for example in relation to anomalous diffusion, but predominantly on lattice-like, annealed[24] and random structures[20,23,25–28]. These limitations leave a fundamental question open, with important applications for the modelling of temporal networks: what are the effects of complex patterns, simultaneously in time and in structure, on the approach to equilibrium of diffusion processes?

To comprehend the interplay between temporal and structural patterns, we focus on a broad class of linear multi-agent systems describing N interacting nodes, defined by

$$D\mathbf{x} = L\mathbf{x} \qquad (1)$$

where x_i, the ith component of $\mathbf{x}$, represents the observed state of node i. The $N \times N$ real matrix L encodes the mutual influences in the network, with non-zero entries indicating the presence of a link. D is either d/dt, the delay $Dx_i(t) = x_i(t+1)$, or any other causal operator acting linearly on the trajectory of each entry $x_i(t)$. This equation couples network structure (represented by L) and time evolution (represented by D) by describing a system where every node i has a state $x_i(t) = Fu_i(t)$, where $u_i(t) = \sum_j L_{ij}x_j(t)$ represents the input, or influence of the neighbouring states on node i. The operator F is the so-called transfer function[29,30], defined as the inverse of D (See Methods).

A classic example is heat diffusion on networks, where every node has a temperature x_i evolving according to the Fourier Law

$$\mu\frac{d\mathbf{x}}{dt} = L\mathbf{x}, \qquad (2)$$

where μ is the characteristic time of the dynamics and L is a Laplacian of the network. The same set of equations can represent, possibly up to a change of variables, a basic model for the evolution of people's opinions[31], robots' positions in the physical space[30,32], approach to synchronisation[33,34] or the dynamics of a continuous-time random walker[35]—our main example from now. In any case, the dynamical properties of the system are described by the spectral properties of the coupling matrix. The constraints imposed by the conservation of probability imply that the Laplacian dynamics is characterised by a stationary state, associated to the dominant eigenvector of L, which we will assume to be unique, as is the case in a large class of systems, for example strongly connected networks. A key quantity is thus the second dominant eigenvalue, also called the spectral gap[6], which provides us with the relaxation time to stationarity, usually called mixing time[36] for stochastic diffusion processes. The spectral gap determines the speed of convergence to the stationary state, and measures the effective size of the system

---

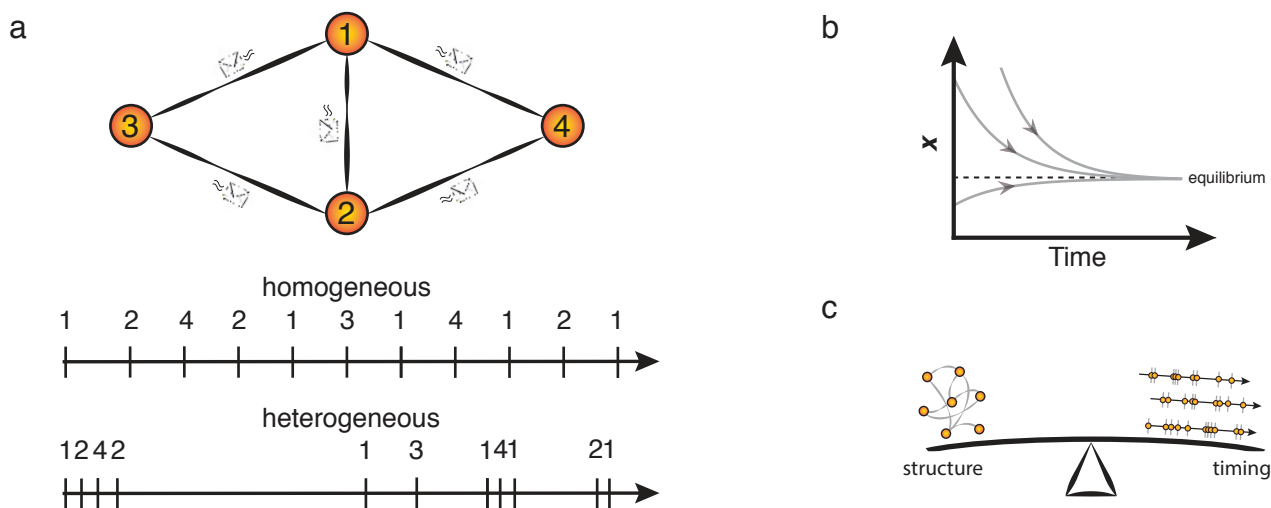*Electronic address: Jean-Charles.Delvenne@uclouvain.be

Figure 1: **Diffusion on temporal networks.** (a) A random walk process can be illustrated by a letter (or banknote, etc.) being randomly passed from neighbour to neighbour on a social network. Temporal patterns of waiting times between arrival and departure of the letter may be homogeneous (for instance discrete or exponentially distributed times), or heterogeneous (for instance bursts). (b) The relaxation time measures the characteristic time to reach equilibrium from any starting condition. (c) The competition between structure and temporal patterns regulates the relaxation time, or mixing time, of stochastic processes.

in terms of dynamics. The spectral gap is also related to important structural and dynamical properties of the system, such as the existence of bottlenecks and communities in the underlying network[7,30].

In this paper, we generalise the concept of spectral gap and of mixing time to random processes with general causal operator D, and focus in detail on operators with long-term memory, naturally emerging in diffusion with bursty dynamics. After showing connections between the theory of random walks and that of multi-agents systems, we identify the temporal and structural mechanisms driving the asymptotic dynamics of the system, and provide examples when each mechanism prevails. By doing so, we show that the form of the temporal operator D may either slow-down or accelerate mixing as compared to the differential operator equation (2). The results are further exploited to assess the possibility of coarse-graining the dynamics based on the network community structure, and tested using numerical simulations on synthetic temporal networks calibrated with empirical data.

Results

Random walks with arbitrary waiting times. The generalised dynamics of the random walker, illustrated in Fig. 1a, is defined as follows. Assume that a walker arrives at node i and hops at a time between Δt and $\Delta t + \delta$ after its arrival. A walker arriving at a node i jumps towards a neighbouring node within a time interval $[\Delta t, \Delta t + \delta]$ with probability $\rho(\Delta t)\delta$ (for small δ). In line with a standard discrete time random walk process, the jump is directed towards a neighbour j with probability P_{ij}. The probability density function $\rho(\Delta t)$ is called the waiting time distribution of the walker. At each hop, the waiting time Δt is reset to zero, and consecutive waiting times are independent. The evolution of probability of presence of a stochastic process in each state is ruled by the so-called master equation, or Kolmogorov forward equation, well known for this family of generalised walks[24,37,38]. We prefer to adopt here the equiv-

alent, dual, viewpoint of Kolmogorov backward equation[39], which belongs to the class of processes defined by equation (1) and can be analysed using the toolbox typical to multi-agent systems[30,32], such as the transfer function formalism and eigenmode decomposition.

We assign a fixed real-valued observable $x_{\text{obs},i}$ to every node i, and consider the value observed at time t by a walker starting initially from node j. This value is a random variable, with an expected value denoted by $x_j(t)$, taking initial value $x_j(0) = x_{\text{obs},j}$. The random walk is ergodic and mixing on a strongly connected aperiodic network (aperiodicity is only required for the discrete-time, when $\rho(\Delta t)$ is a Dirac distribution). In this case, $\mathbf{x}(t)$ describes a consensus dynamics, meaning that the individual values asymptotically converge to one another, $x_i(\infty) = x_j(\infty)$. If we choose the observable $x_{\text{obs},j} = 0$ at all nodes j, except 1 at node k, then $x_i(t)$ is precisely the probability for the random walker starting at i to be found at k at time t. Therefore the Kolmogorov backward equation, a consensus dynamics, embeds in particular the evolution of the probability of presence on nodes.

The walker, assumed to have just hopped at time zero and finding itself at node i (another origin of time would not change the asymptotic decay times, of main interest in this work), hops again at time Δt with probability density $\rho(\Delta t)$, and moves towards a neighbour j with probability P_{ij}. The expected value observed by the walker at time t is $x_{\text{obs},i}$ if it is still waiting for its first hop ($t < \Delta t$) and otherwise $\sum_j P_{ij} x_j(t - \Delta t)$, by induction on the number of hops. Therefore we obtain the vector equation

$$\mathbf{x}(t) = \mathbf{x}_{\text{obs}} \int_t^\infty \rho(\Delta t)\mathrm{d}\Delta t + \int_0^t P\mathbf{x}(t - \Delta t)\rho(\Delta t)\mathrm{d}\Delta t, \quad (3)$$

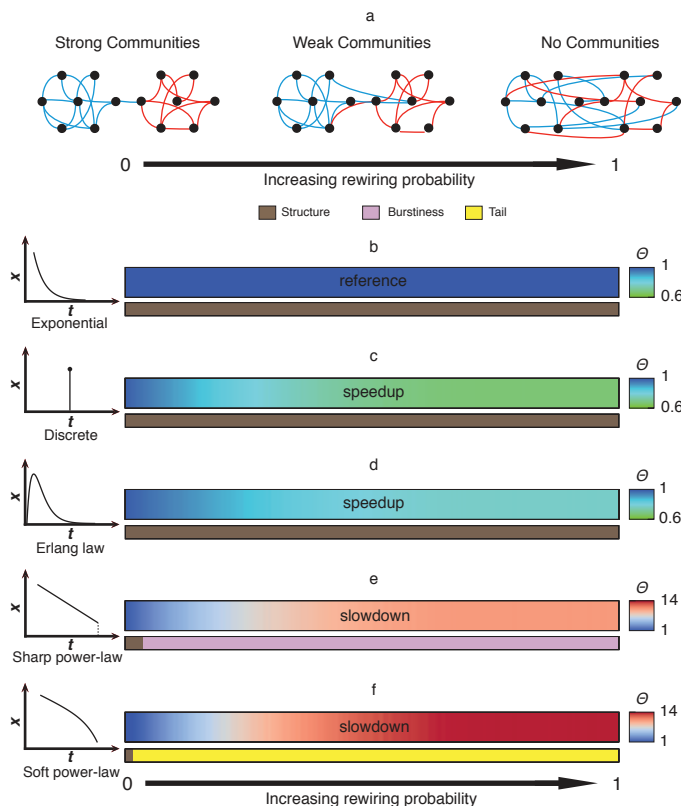with the discrete transition matrix P of entries P_{ij}. The con-

Figure 2: **Network structure or waiting times regulate mixing times.** We illustrate the findings simple network composed of two communities, where the community structure is modulated by rewiring edges starting from two separated random graphs, as shown in a). Each panel (b-f) shows (on the top bar) for each value of the rewiring probability the slowdown factor Θ, ratio of the exact mixing time for a given waiting time distribution and for the exponential case (Poisson process). Each panel corresponds to a different waiting time distribution and the colorbar scale on the right represents the values of Θ for the respective colormap. Θ close to one indicates that non-trivial waiting time patterns can be neglected in the determination of the mixing time. On the other hand, a high Θ indicates that the details of the network structure are overshadowed by the temporal dynamics. b) shows the reference case, that is, the exponential waiting times. We observe a speedup ($\Theta < 1$) for c) Dirac (discrete) and d) Erlang distributions and a slow-down ($\Theta > 1$) for e) a power-law with sharp cutoff and for f) a power-law with soft cutoff. From left to right, we show how this factor depends on the network structure by progressively removing the community structure (increasing ϵ). The bottom bar of each panel shows, according to equation (14), which is the dominant factor, either structure, burstiness, or tail, regulating the relaxation time for each model of waiting times. See Methods for details on the synthetic networks and waiting time distributions.

volution in time in the last term calls for a Laplace transform

$$\mathbf{x}(s) = \int_0^\infty \mathbf{x}(t)\, e^{-st} \mathrm{d}t. \qquad (4)$$

For simplicity, we use the same notations for functions in the time and in the Laplace domain, only distinguished by their variable, namely t or Δt for time, and s for Laplace. This is justified as time and Laplace domain representations encode the same single physical object, for example a probability or an observable. The same holds for operator D, thought of as acting in the time domain (for instance $D = \mathrm{d}/\mathrm{d}t$) or the Laplace domain ($D = s$) according to the context.

The Laplace transform $\rho(s)$ is the moment generating function of the waiting time distribution $\rho(\Delta t)$, as it encodes the moment in its Taylor series $\rho(s) = 1 - \mu s + (\mu^2 + \sigma^2)\frac{s^2}{2} - \cdots$, where μ is the expected waiting time (first moment), σ^2 is the variance and $\mu^2 + \sigma^2$ is the second moment. Using the fact that convolution (respectively, integration from 0 to t) in the time domain corresponds to the usual product (respectively, division by s) in the Laplace domain[40], equation (3) reduces to

$$\mathbf{x}(s) = \frac{1 - \rho(s)}{s}\mathbf{x}_{\mathrm{obs}} + \rho(s)P\mathbf{x}(s), \qquad (5)$$

or equivalently

$$\left(\frac{1}{\rho(s)} - 1\right)\mathbf{x}(s) = \left(\frac{1}{\rho(s)} - 1\right)\frac{1}{s}\mathbf{x}(t=0) + L\mathbf{x}(s), \qquad (6)$$

where we have made the dependence on the initial condition explicit by using the relation $\mathbf{x}_{\mathrm{obs}} = \mathbf{x}(t = 0)$ and where $L = P - I$ denotes the (normalised) Laplacian of the network. This is an instance of equation (1), which shows that an input-output relationship is often best expressed in the Laplace domain rather than in the temporal domain. In this case, the transfer function $F(s)$ is defined by the algebraic relation $F^{-1}(s) = 1/\rho(s) - 1 = D(s)$, up to the initial condition term, implicit in equation (1). See Methods for a derivation of equation (1) in a more general context.

From temporal networks to random walks. Diffusive processes often take place on temporal networks where individuals initiate from time to time short-lived contacts with their neighbours. A random walker can represent for example a letter or a banknote passing from hand to hand through first contact initiated by the current node. The formalism described above focuses on the statistical properties of the waiting times of a walker on a node[15,25], and not of the inter-contact times (the times between two subsequent contacts from a given node to another), as often considered in the literature[20,27,28,42,43,47]. To illustrate this difference, let us consider an idealised scenario, where the network looks locally like a directed tree, in order to avoid indirect correlations due to cycles[44], where the inter-contact time t between two contacts initiated by the same node is characterised by the same probability distribution $\rho_{\mathrm{contact}}(t)$, and where activations on different edges are an independent random process. The corresponding waiting time distribution $\rho(\Delta t)$ for the random walker can be determined from ρ_{contact}[45,46].

For example, the classic inspection paradox, or bus paradox, observes that the waiting time has a mean

$$\mu = 1/2\mu_{\mathrm{contact}}(1 + \sigma_{\mathrm{contact}}^2/\mu_{\mathrm{contact}}^2) \qquad (7)$$

which can be much higher than the average inter-contact time μ_{contact} in case of bursts. This fact has been used in the literature to deduce that burst contact statistics slow down diffusion in a complex network[47]. Nevertheless it should not be confounded with the results presented in this paper, which will focus on the statistical properties of the waiting time distribution and identify, amongst others, that its variance plays a significant role on the asymptotic behaviour of

the walker. In the above scenario, the variance of $\rho(\Delta t)$ depends on the third moment of the inter-contact time distribution $\rho_{\text{contact}}(t)$, and is associated to a mechanism distinct from that of the bus paradox (7).

Eigenmodes for heterogeneous temporal operators.
Equation (1) typically takes the form of an integro-differential equation. However, it simplifies into the differential equation (2) in the case of a memoryless random walker. Memoryless refers to the case when the probability of hopping between Δt and $\Delta t + \delta$, knowing that the walker has waited at least Δt, is independent of Δt. This leads to an exponential waiting time distribution[45] $\rho(\Delta t) = e^{-\Delta t/\mu}/\mu$, in other words an unconditional probability $e^{-\Delta t/\mu}\delta/\mu$ of jumping between Δt and $\Delta t + \delta$, for small δ and for mean waiting time μ. In that case, we find $D(s) = 1/\rho(s) - 1 = \mu s$ in the Laplace domain, indeed recovering equation (2) in the time domain. The differential equation can then be analysed by changing the variables $\mathbf{x}$ to a linear combination of the eigenvectors $\mathbf{v}_k$ of the Laplacian L, of eigenvalues $\lambda_0 = 0 > \lambda_1, \lambda_2, \ldots, \lambda_{N-1} \geq -2$ as follows[30]:

$$\mathbf{x}(t) = \sum_k z_k(t)\mathbf{v}_k. \tag{8}$$

For simplicity we have supposed that the underlying network is undirected, connected and, in case of a discrete-time random walk, non-bipartite. Thus the eigenvalues are real and the stationary state is uniquely defined[6]. Every $z_k(t)$, solution of $Dz_k(t) = \lambda_k z_k(t)$, is called an eigenfunction of the operator D, and here takes the form of a decaying exponential. The problem is thus solved by decoupling structural and temporal variables, first by identifying structural eigenvectors $(\mathbf{v}_k)$, and then how they evolve in time $(z_k(t))$. The resulting fundamental solutions $z_k(t)\mathbf{v}_k$ for equation (1) are called modes, or eigenmodes, of the system.

A similar analysis can be performed in the Laplace domain in the case of an arbitrary waiting time distribution $\rho(\Delta t)$, and thus when equation (1) has an arbitrary temporal operator D. In that case, the elementary solutions $z_k(t)$, associated to $z_k(s)$, need not be exponential functions, and are obtained as solutions of equation (6), where the Laplacian is reduced to its eigenvalue λ_k

$$z_k(s) = \frac{1}{1 - \frac{\lambda_k}{1/\rho(s)-1}} \frac{z_k(t=0)}{s}, \tag{9}$$

As before, any trajectory of the system can be expressed as a linear superposition of some or all the N modes.

Characteristic decay times of eigenmodes. Despite their non-exponential nature, a broad class of eigenfunctions (9) still have a characteristic time τ_k describing the asymptotic decay of $z_k(t)$ to equilibrium as e^{-t/τ_k} (Fig. 1b). We find that the decay time can be accurately estimated by performing a small s expansion in the Laplace domain (see Methods for the derivation and range of validity):

$$\tau_k \approx \mu(|\lambda_k|^{-1} + \beta), \tag{10}$$

Figure 3: **Dominating mechanisms on temporal networks.** (a) shows the regulating potential of each mechanism according to equation (14), either structure ($\mu\epsilon^{-1}$), burstiness ($\mu\beta$), or tail (τ_{tail}), for the synthetic networks based on the empirical network structures and on the waiting times obtained by simulating a random walk on these structures. Either structure (brown) or tail (yellow) dominates the mixing time, also when structure or contact times are randomized independently (b-d).

where

$$\beta = \frac{\sigma^2 - \mu^2}{2\mu^2} \tag{11}$$

is a measure of the burstiness of the temporal process based on the first and second moments of its distribution. Burstiness β equals to zero for a Poisson process (memoryless waiting times, $\rho(\Delta t) = e^{-\Delta t/\mu}/\mu$) and ranges from $-1/2$ (if $\rho(\Delta t)$ is a Dirac distribution) to arbitrarily large positive values (for highly bursty activity). This expression emerges naturally from the dynamical process and can be viewed as a measure of burstiness, equivalent to the commonly used burstiness measure[48].

The estimate (10) can be further tightened whenever the distribution $\rho(\Delta t)$ of waiting times contains a fat tail, possibly softened by an exponential tail. The archetypical example is a power-law with soft cut-off, $\rho(\Delta t) \propto (\Delta t + A)^{-\gamma}e^{-\Delta t/\tau_{\text{tail}}}$, a frequent model in human dynamics supported by empirical evidence[9,14,42]. More generally $\rho(\Delta t) \propto \rho_0(\Delta t)e^{-t/\tau_{\text{tail}}}$, where ρ_0 is a fat tailed distribution, that is, decreasing subexponentially. The non-analytic point created in $\rho(s)$ by the fat tail leads to an additional term in the characteristic time (see Methods)

$$\tau_k \approx \max(\mu(|\lambda_k|^{-1} + \beta), \tau_{\text{tail}}). \tag{12}$$

which can be approximated as follows, if $|\lambda_k|^{-1}$ and β have different orders of magnitude

$$\tau_k \approx \max(\mu|\lambda_k|^{-1}, \mu\beta, \tau_{\text{tail}}). \tag{13}$$

Mixing times and dominating mechanisms. Of partic-

ular importance is the slowest non-stationary mode ($k = 1$), or mixing mode, the characteristic decay time of which represents the worst-case relaxation time of any initial condition to stationarity. We call this time the mixing time of the process, in generalisation of the classic memoryless case, where it is given by $\tau_{\text{mix}} = \mu\epsilon^{-1}$, determined by the so-called spectral gap $\epsilon = -\lambda_1 > 0$[6,36]. This quantity is related to the presence of bottlenecks (that is, weak connections between groups of highly connected nodes, a.k.a. network communities) in the network (via Cheeger's inequality[49,50]). It is approximated as

$$\tau_{\text{mix}} \approx \max(\mu\epsilon^{-1}, \mu\beta, \tau_{\text{tail}}). \tag{14}$$

This expression shows that the asymptotic dynamics of each mode is determined by the competition between three factors: a structural factor ($\mu\epsilon^{-1}$), associated to the spectral properties of the Laplacian of the underlying network, and two temporal factors associated to the shape of the waiting time distribution, namely its burstiness ($\mu\beta$) and its exponential tail (τ_{tail}). The slowest (largest) of these factors dominates the asymptotic dynamics (Fig. 1c).

We emphasise that the burstiness and the fat tail effects are not necessarily related[46]. For instance a power-law $\rho(\Delta t) \propto (\Delta t + 1)^{-3}$ restricted to times $\Delta t \leq T$ (sharp cutoff), is arbitrarily bursty but has no tail at all ($\tau_{\text{tail}} = 0$) for large but finite T. On the other hand, the delayed power-law $\rho(\Delta t) \propto (\Delta t - T)^{-4}$, restricted to times $\Delta t \geq T + 1$ has a fat tail (it decreases subexponentially, $\tau_{\text{tail}} = \infty$) but low burstiness β for large T, as the mean time μ increases without bound and the variance remains constant. In the latter case, as for all pure power laws, $\rho(\Delta t) \sim \Delta t^{-\gamma}$ for large Δt, the mixing time τ_{mix}, as τ_{tail}, is actually infinite, reflecting that mixing, or relaxation to stationarity, occurs only with subexponential convergence. In general, the properties of the tail of the distribution depend on the high-order moments of the distribution, and not only on the first two moments as captured in the coefficient of burstiness.

In Fig. 2, we study the mechanism dominating the mixing time in toy synthetic temporal networks with different waiting time distributions. Not surprisingly, structure is the driving mechanism when waiting times are narrowly distributed around a mean value, as in the case of Dirac (discrete-time, Fig. 2c) and Erlang distributions (resembling a discrete-time distribution with small fluctuations, Fig. 2d). For those, the slowdown factor $\Theta = \tau_{\text{mix}}/(\mu\epsilon^{-1})$, comparing the exact mixing time (computed with equation (18) in Methods), to what it would be with memoryless waiting times ($\mu\epsilon^{-1}$), takes value in $[0.6, 1]$. This indicates a limited speed up of the mixing due to negative burstiness, $\beta < 0$, while the structure plays a major role through ϵ traversing orders of magnitude.

On the other hand, competition between structure and time appears in scenarios of high temporal heterogeneity. For example, only strong communities are able to dominate power-law waiting times (Fig. 2e-f). The effect of structure on the mixing times is otherwise removed as burstiness (Fig. 2e) or tail (Fig. 2f) becomes the leading mechanism, scaling the mixing times up to 14-fold in the shown configurations. The transitions between the different mechanisms

for a range of power-law configurations are presented in Figure 6 and Supplementary Note 1.

Model reduction. The use of coarse-graining through time scale separation, which is the separate treatment of fast and slow dynamics that coexist inside a system, is crucial to reduce the complexity of systems made of a large number of interacting entities[51-53]. This procedure is well known for differential equations like (2). In this case, it consists in neglecting fast decaying modes, for example with decay time less than a certain threshold τ_{tresh}—an approximation invalid for early times but acceptable for times significantly larger than τ_{tresh}. Only the dominant modes, thus fewer variables, are left in the reduced model. Decreasing the threshold time, one produces a full hierarchy of increasingly more accurate models, but also with increasingly more variables. Reduced models have a clear interpretation from the structural point of view, as fast modes typically correspond to the fast convergence of the probabilities of presence on nodes to a quasi-equilibrium within a network community. This process is followed by a slow equilibration of the population of random walkers trapped in each community to a global equilibrium[53-55]. Each new reduced variable can therefore be interpreted as the slow-varying probability of presence in each community, as if it had been collapsed into a single node. The hierarchy of increasingly more accurate and complex reduced models corresponds in this structural picture to a hierarchy of increasingly finer partitions into communities. Given that the decay times of the different modes in equation (2) correspond to the Laplacian eigenvalues, the k-community partition is unsurprisingly found to be encoded into the k dominant Laplacian eigenvectors, in a way that is decoded by spectral clustering algorithms[56,57].

This methodology can be extended to more general temporal operators D, where the successive decay times now depend on the Laplacian eigenvalues and the temporal characteristics, following equation (13). However, the effect of temporality may have non-trivial consequences, as it may limit the number of reduced models. As an extreme scenario, let us consider a complex network in which the transient dynamics is dominated by the fat tail of the waiting time distribution, such that all transient modes decay for large times as $e^{-t/\tau_{\text{tail}}}$. Following equation (8), the approach to stationarity is described by the superposition of modes, all of the form $f_k(t)e^{-t/\tau_{\text{tail}}}\mathbf{v}_k$, for various functions $f_k(t)$ that decay more slowly than exponentially, as for example $f_k(t) \propto t^{-\gamma_k}$. Among those, any mode can only become negligible with respect to another after a very long time. For all practical purposes, the system exhibits an inherently complex dynamics with no time scale separation, as the fine scale structure (the finest being at the level of a single node) and the large scale network communities have equal or similar impact on the random walk dynamics. As a consequence, only two reduced models are available, one in which all N modes are considered, and another described solely by the stationary state.

In general, only the Laplacian eigenvalues smaller in magnitude than μ/τ_{tail} determine corresponding decay times. Just like in classical memoryless case, they generate a hierarchy of times scales and reduced models corresponding
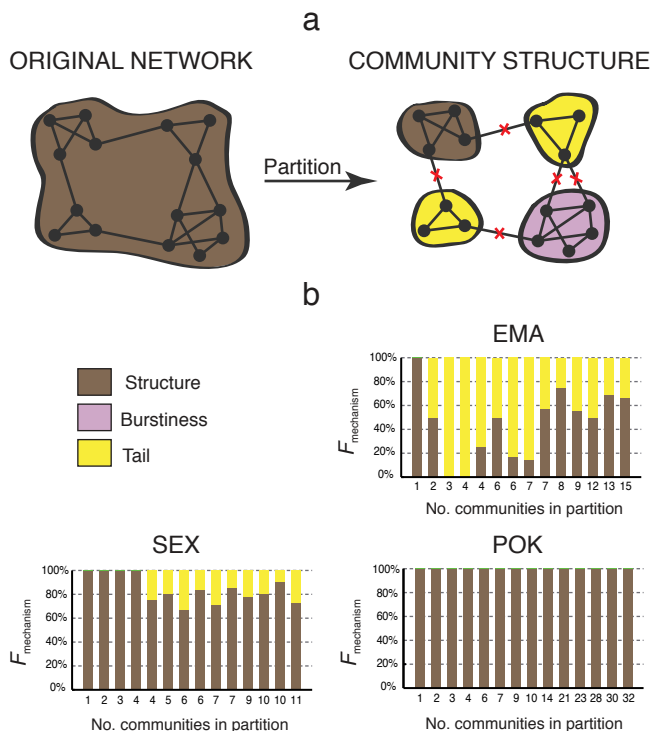
Figure 4: **Global mixing time vs mixing times in communities.** (a) While the mixing time of diffusion on a network may be defined by the community structure, diffusion in the communities themselves may be regulated by structural or temporal patterns. (b) The original network constructed from an empirical dataset is divided into communities with a multi-resolution community detection algorithm, for different values of the resolution parameter (See Methods). For each partition, we identify the fraction of communities larger than 10 nodes (y-axis) where either structure or tail dominates (according to equation (14)), and plot against the number of communities detected at each partition (x-axis). This confirms our spectral analysis, that predicts a compact description of the diffusion dynamics on EMA into three modes describing a probability flow between three communities, each dominated by fat tailed waiting times. Also in agreement with the spectral viewpoint, this is not the case for SEX or POK, unless at the level of a few-node communities.

to multiple levels of community structures. For instance, if 5 eigenvalues are smaller than μ/τ_{tail}, thus $0 = \lambda_0 > \lambda_1 > \cdots > \lambda_4 > -\mu/\tau_{\text{tail}}$, then there is a reduced diffusion model capturing the movement of the random walker between 5 aggregated nodes. The mixing, internal to each community and decaying as $e^{-t/\tau_{\text{tail}}}$, is considered instantaneous at large enough time scales. Coarser aggregation, for example based on two communities and two eigenvalues λ_0, λ_1, may be relevant although valid only at even larger times. However, aggregation based on any finer partitioning (other than into one-node communities) has the same accuracy as the five-community model, and thus little practical value as a reduced model. The degeneracy of characteristic decay times therefore limits the number of useful reduced models. This implies that a decomposition into communities is not necessarily associated to a time scale separation, or a reduced model, of the dynamics. For this reason, only models incorporating 1 to 5 or N dominating modes are adequate for this particular example. This smaller choice of reduced models has ambivalent consequences. It limits the set of resolutions at which to describe the dynamics, but also provides a natural level for community structure (an open problem in multi-scale community detection[58–62]), defined

as the finest partitioning yielding a reduced model for the system.

Numerical Analysis. Equation (14) shows that the mixing time depends in first approximation only on the mean, variance and tail of the waiting time distribution, while other properties of the distribution are irrelevant. With numerical simulations, we validate this approximation and provide some quantitative intuition on the competition between the three factors regulating the mixing. We construct synthetic temporal networks respecting our assumptions, for example stationarity or absence of correlations, and calibrate them with the static structure and the inter-contact time distribution observed on a number of empirical datasets. The empirical networks used for calibration correspond to face-to-face interactions between visitors in a museum (SPM), between conference attendees (SPC)[13], and between hospital staff (SPH)[16]; email communication within a university (EMA)[10]; sexual contacts between sex-workers and their clients (SEX)[14]; and communication between members of a dating site (POK)[12] (See Methods). On these networks we observe the waiting time distribution and the spectral gap, which allows to compute both the exact mixing time, with equation (18) in Methods, and its approximation in equation (14). The results, reported in Table 1, show a good agreement, except for SPC, as analysed later in this section.

	SPM	SPC	SPH	EMA	SEX	POK
Θ	1.77	1.81	1.54	1.19	1.02	1.02
Exact Mixing	54	205	671	718	7,887	29,347
Approximate Mixing	61	313	664	603	7,741	28,685
Relative Difference (%)	13.0	52.7	-1.0	-16	-1.9	-2.3

Table 1: **Exact vs. approximate mixing times.** The table shows that for six configurations of temporal networks the slowdown factor $\Theta = \tau_{\text{mix}}/(\mu\epsilon^{-1})$, that is the ratio of the exact mixing time, calculated from equation (18) using the simulated waiting time distribution, to the mixing time we would have for the exponential case (Poisson process, with same mean). The results indicate strong (SPM, SPC and SPH), medium (EMA) and weak (SEX and POK) slowdown ($\Theta > 1$). The approximate mixing time, computed with equation (14), shows a good agreement with the exact value.

Except for SEX[43] and POK datasets, in the other cases the temporal heterogeneity substantially increases the mixing times (see slowdown factor Θ in Table 1). Fig. 3 shows that in these networks the dominant factor regulating the mixing time depends on the characteristics of the system. Fat tails of the waiting time distribution drive the relaxation for the cases of face-to-face contacts (SPM, SPC, SPH). Structure however is the leading mechanism behind the networks corresponding to other situations of human communication (EMA, SEX, POK).

If communities are completely removed by randomising the network structure, the link sparsity of SEX and POK networks guarantees that structure remains dominating, as the sparsity results in the inevitable creation of bottlenecks for diffusion even in a random network. On the other hand, in the EMA network, which has a relatively dense connected structure, absence of communities leads to temporal dominance (Fig. 3). Finally, when the contact times are uniformly distributed, we recover the well-known result that network

structure, with or without communities (Fig. 3), is the main factor regulating the convergence to stationarity.

Because the raw empirical temporal networks do not necessarily abide by our simplifying assumptions, one cannot validate our theoretical derivations, for example, by estimating the mixing time directly from simulations on empirical data. Indeed the periodic rythms or correlations between successive jumps or other temporal patterns may induce effects not captured by our formula, as discussed in Table 3 and Supplementary Note 3.

Empirical data are collected during a finite time span T, leaving the choice between a model for $\rho(\Delta t)$ that is limited to T (sharp cut-off) or extrapolated to infinite times with a tail decay (soft cut-off). The latter may occur for instance if we have theoretical or practical reasons to prefer a fat-tail-based model (for instance power-law) with exponential cut-off for $\rho(\Delta t)$. This choice is of little impact on the results, provided a sufficient observation time T (Supplementary Note 2). When temporal patterns trump structure, we have a self consistency condition $\tau_{\mathrm{mix}} \approx \mu(\epsilon^{-1} + \beta) \approx \tau_{\mathrm{tail}}$ (Fig. 3) for SPM and SPH), expressing that both models lead to similar mixing times, albeit through the different mechanisms of either burstiness (sharp) or fat-tail (soft). For SPC, self consistency is not attained. This happens possibly because the observation time of ~ 2 days is not sufficiently large to dilute the irregularity on the observed inter-contact time distribution (thus in the simulated waiting times) induced by the inactivity during night periods. Consequently, this makes the tail difficult to measure as it has not clearly emerged from a still transient behaviour, and a significant mismatch is observed between exact and approximate mixing (Fig. 3).

We evaluate the possible sizes of reduced models for the data as follows. As the eigenvalues λ_k of the Laplacian range between 0 and -2, the number of modes with structurally determined decay times ($-\mu/\tau_{\mathrm{tail}} < \lambda_k \leq 0$) can be roughly evaluated to $N\mu/2\tau_{\mathrm{tail}}$ on an N-node network, if eigenvalues are evenly spaced. A similar reasoning holds whenever burstiness dominates the tail effect. This analysis reveals three different scenarios in the structure-dominated datasets considered above: i) SEX benefits from a full hierarchy of reduced models, as $\mu/2\tau_{\mathrm{tail}} > 1$. ii) The dynamics on EMA can be approximated with just three modes, associated to three network communities, as $0 = \lambda_0 > \lambda_1 > \lambda_2 > -\mu/\tau_{\mathrm{tail}} > \lambda_3 > \ldots$, while a more detailed description, unless at the node level, would not gain any significant accuracy as further modes are all degenerate with the same decay time μ/τ_{tail}. iii) POK exhibits a practically full hierarchy of reduced models as around one third of its modes are determined by structure, $\mu/2\tau_{\mathrm{tail}} = 0.33$, therefore the finest reduced model is at a few-node community level. This spectral analysis is comforted by applying a multiscale community detection algorithm to the empirical networks, which finds a partition into three communities dominated by fat-tail effects for EMA, but not for POK or SEX (Fig. 4).

Discussion

We have presented a unified mathematical framework to calculate the relaxation time to equilibrium in a wide variety of stochastic processes on networked temporal systems. Our

formalism is able to refer to arbitrary linear multi-agent complex systems, including linearisations of non-linear dynamical models such as Kuramoto oscillators or non-Laplacian diffusion dynamics such as SIR-like epidemics, as detailed in Methods. It is also possible to accommodate non-uniformity of parameters as nodes are not identical in real systems. Our results are particularly relevant to improve the understanding of temporal networks, by highlighting the important interplay between structure and the temporal statistics of the network. Our formalism is different from previous studies of random walks on temporal networks that focus on homogeneous temporal patterns[23] or do not account for the competition between structure and time[15]. We emphasise that questions related to ordering, not timing of events, such as the number of hops before stationarity or the succession of nodes most probably traversed by the walker, depend on the structure alone and not of hops timing statistics. Moreover, key mechanisms have been left aside in our modelling approach, such as the non-stationarity or periodicity of most empirical networks[19–21], and the existence of correlations between edge activations and therefore preferred pathways of diffusion[17,18,42]. Important future work includes their incorporation in our mathematical framework, and the identification of dominant mechanisms in empirical data.

We have shown that, in the absence of some temporal correlations, the characteristic times of the dynamics are dominated either by temporal or by structural heterogeneities, as those observed in real-life systems. The competing factors are not only observed in different classes of networks modelled from empirical data but also at different hierarchical levels of the same network represented by its different scales of community structure. We have also identified two contrasting metrics of the statistics of waiting times, burstiness and fat-tails. We have shown that they regulate the dynamics on the network in different ways. In systems where temporal patterns are the dominating factor, the reduced models obtained by aggregation of communities as commonly used in practice are not necessarily relevant, as small scale details are impenetrably intertwined with large scale structure to form a complex global dynamics. In general, the temporal characteristics impose the natural description levels of the dynamics. Altogether, our results suggest the need for a critical assessment of a complexity/accuracy trade-off when modelling network dynamics. In some classes of real-world systems, the burden of increased model complexity may not compensate the incremental gain in knowledge, while other systems require the fine network structure as a key ingredient to any realistic modelling.

Methods

Overview. We provide a description of multi-agent linear dynamics, detail the derivation of approximation (12) and its range of validity, illustrate the generality of the framework beyond random walks, and describe the empirical data and numerical implementations. In the Supplementary Information, we discuss on power-laws and cut-offs (Supplementary Note 1 and Figure 6), the evaluation of waiting time distributions in empirical data (Supplementary Note 2), and the adequacy of our theoretical framework in empirical networks with correlations and non-stationarity (Supplementary Note 3).

Linear multi-agent systems. We derive equation (1) for general linear networked systems, or multi-agent systems[30], with an illustration on consensus dynamics. Every node i is modelled by a linear agent

whose internal state is initially zero and that converts an input signal u_i with an operator F_i—the so-called transfer function—into a state signal $x_i = F_i u_i$. Here, u_i represents a time trajectory $u_i(t)$ or its Laplace transform $u_i(s)$, and similarly for the state trajectory x_i. The transfer function F_i is a operator mapping input trajectories u_i to state trajectories x_i, that is requested to be linear, causal (if $u_i(t) = 0$ for all $t < T$, then the same holds for $x_i(t)$), and time invariant (shifting $u_i(t)$ in time results in shifting $x_i(t)$). Under those conditions, this operator takes the form $x_i(s) = F_i(s)u_i(s)$, a simple product of functions, in the Laplace domain.

When using Laplace transforms it is customary to use time domain trajectories that are zero for negative times. We account for the state initial condition $x_i(0) = x_{i,0}$ with another input term v_i that can be for example an impulse exciting the rest state to any desired initial condition at time zero. The agent dynamics therefore writes $x_i(s) = F_i(s)(u_i(s) + v_i(s))$. For instance if F_i is the integration operator in the time domain, $x_i(t) = x_{i,0} + \int_0^t u_i(t')dt'$ from the initial value at time 0, or in other terms $x_i(t) = \int_0^t u_i(t') + x_{i,0}\delta(t')dt'$ (for the Dirac impulse $\delta(t)$) then in the Laplace domain we find $x_i(s) = 1/s(u_i(s) + x_{i,0})$; in this case the Laplace domain operator is the multiplication by $1/s$, and v_i is a Dirac impulse in the time domain, a constant in the Laplace domain that encodes the initial condition. One can as well invert the transfer operator, $D_i = F_i^{-1}$, and write $dx_i/dt = u_i(t) + x_{i,0}\delta(t)$, or $sx_i(s) = u_i(s) + x_{i,0}$; here the Dirac impulse is seen as arising from differentiating the discontinuous step of the state from rest to initial condition $x_{i,0}$.

Connecting agents such that the input $u_i(t)$ of agent i is a weighted sum of other agents' states $\sum_j L_{ij}x_j(t)$ leads to a global dynamics $d\mathbf{x}/dt = L\mathbf{x} + \mathbf{x_0}\delta(t)$, or equivalently $s\mathbf{x}(s) = L\mathbf{x}(s) + \mathbf{x_0}$, where L is the matrix of entries L_{ij}, $\mathbf{x}$ is the vector of states x_i, and $\mathbf{x_0}$ the vector of initial conditions $x_{i,0}$. When L is a Laplacian (rows summing to zero, nonnegative off-diagonal weights), this is a simple consensus system where agents (for instance robots or individuals) change their state (for instance position, opinion modelled as a real number) towards an average of their neighbours' states. With arbitrary interconnection weights and arbitrary transfer functions, one similarly obtains

$$D\mathbf{x} = L\mathbf{x} + \mathbf{v} \tag{15}$$

where D is the diagonal matrix of transfer functions F_i^{-1} (D reduces to a scalar in case of identical agents $F_i = F_j$), and $\mathbf{v}$ the vector of initial condition input v_i. The case when L remains Laplacian, but D is arbitrary, describes general, higher-order, consensus systems where the agents converge to equal values on a strongly connected network through arbitrarily complicated internal dynamics, for example modelling realistic robots or vehicles. For example, vehicles of mass m driven by a force and friction will obey $m\ddot{x} + \alpha\dot{x} = L\mathbf{x}$. An abuse of notation allows to drop implicitly the initial condition and write simply $d\mathbf{x}/dt = L\mathbf{x}$, and more generally equation (1), $D\mathbf{x} = L\mathbf{x}$, instead of equation (15) above. Discrete-time systems are recovered with a discrete-derivative operator $D\mathbf{x}(t) = \mathbf{x}(t+1) - \mathbf{x}(t)$.

Exact and approximate characteristic times for random walkers. We derive the approximate characteristic times (12) of decaying modes associated to a general random walk. The dominant mode associated to $\lambda_0 = 0$ is the stationary distribution, unique for an ergodic random walk. The relaxation time to the stationary distribution, generalising the well-known mixing time of Markov chain theory[36], is the characteristic time of the slowest decaying mode after the stationary mode, usually associated to λ_1, as we suppose here. An example of a case when λ_1 is not associated to the mixing mode, but rather λ_N, is the discrete-random walk, where the distribution is a Dirac distribution at μ, with zero variance, and the network is bipartite or close to bipartite.

The characteristic decay time of a time-domain function $f(t) \sim e^{-t/\tau_{\text{decay}}}$ can be found in the Laplace domain, as $f(s)$ is defined and analytic over all s of real part larger than $-1/\tau_{\text{decay}}$, but undefined or non-analytic in at least one point s of real part $-1/\tau_{\text{decay}}$. For example the Laplace transform of $e^{-t/\tau_{\text{decay}}}$ is $1/(1 + \tau_{\text{decay}}s)$, with pole at $s = -1/\tau_{\text{decay}}$, and the Laplace transform of $(t+1)^{-\gamma}e^{-t/\tau_{\text{decay}}}$ for $\gamma > 1$, is defined, but non-analytic, at $s = -1/\tau_{\text{decay}}$.

The eigenfunction $z_k(t)$ associated to λ_k is the solution of $D(s)z_k(s) = \lambda_k z_k(s) + D(s)z_k(t=0)/s$, derived in the text as equation (6), with $D(s) = 1/\rho(s) - 1$ and L replaced with λ_k. The decay

time of $z_k(t)$ is found if we find the right-most non-analytic point of

$$z_k(s) = \frac{1}{D(s) - \lambda_k} \frac{D(s)}{s} z_k(t = 0), \tag{16}$$

derived in the main text as equation (9). Note that for $\lambda_k \neq 0$, the singularity in $s = 0$ is only apparent as $D(s) = \mu s - (\sigma^2 - \mu^2)s^2/2 + \ldots$ This expression is non-analytic in s in exactly two circumstances: i. $D(s)$ is non-meromorphic in s, meaning that it cannot be defined analytically in a punctured neighbourhood of s (a neighbourhood of s, minus s) or ii. $D(s) = \lambda_k$.

Case i. happens when $\rho(s)$ is non-meromorphic in s. A typical example is when $\rho(\Delta t) \propto \rho_0(\Delta t)e^{-\Delta t/\tau_{\text{tail}}}$, where $\rho_0(t)$ is a fat-tailed distribution, in other words decreases more slowly than any exponential, leading to a non-meromorphic transfer function—a rare occurrence in standard systems theory. The non-analytic point is precisely $s = -1/\tau_{\text{tail}}$.

Case ii. commands to solve the equation

$$\rho(s) = 1/(1 + \lambda_k), \tag{17}$$

whose solution s_k enters the expression for the exact characteristic decay time, combining the two cases:

$$\tau_k = \max(-1/s_k, \tau_{\text{tail}}), \tag{18}$$

which can be approximated from an expansion of $\rho(s)$. Note that $\rho(s) = 1 - \mu s + (\mu^2 + \sigma^2)s^2/2 - \cdots$ is also called the moment generating function, whose successive derivatives around $s = 0$ are the moments of the distribution, up to sign. We consider the Padé (that is, rational) approximation:

$$\rho(s) \approx \frac{2\mu + (\sigma^2 - \mu^2)s}{2\mu + (\sigma^2 + \mu^2)s}, \tag{19}$$

equivalent for small s to a second-order Taylor approximation in terms of accuracy. This approximates the transfer function $D^{-1}(s) = F(s) = \rho(s)/(1 - \rho(s))$ as

$$F(s) \approx \frac{1}{\mu s} + \beta, \tag{20}$$

where the adimensional term $\beta = \frac{\sigma^2 - \mu^2}{2\mu^2}$ takes its minimum value at $-1/2$ for the discrete-time random walk, zero for a memoryless process, and arbitrary large values for highly heterogeneous distributions. It is therefore a suitable measure of the burstiness of the process[42,48]. Such a transfer function is called Proportional-Integral, a common class in systems theory[29]. The equation $D(s) = \lambda_k$ is approximately solved by $1/s_k \approx \mu(|\lambda_k|^{-1} + \beta)$. A possible non-analytic point at $s = -1/\tau_{\text{tail}}$ caps the characteristic decay time of eigenfunction k to

$$\tau_k \approx \max(\mu(|\lambda_k|^{-1} + \beta), \tau_{\text{tail}}). \tag{12}$$

If λ_k^{-1} and β are of different orders of magnitude, one may further approximate as

$$\tau_k \approx \max(\mu|\lambda_k|^{-1}, \mu\beta, \tau_{\text{tail}}), \tag{13}$$

in which the influence of the fat tail, the structure and the burstiness are decoupled. The mixing time is associated with λ_1. Fig. 5 shows that typically half of the modes have structurally determined decay times, namely those with positive $1 + \lambda_k$, if eigenvalues are uniformly spaced between 0 and -2, in apparent contradiction with the back of the envelope calculation $N\mu/2\tau_{\text{tail}}$ proposed in the main text, which can reach N. This is a consequence of the progressive degradation of the Padé approximation for larger s (larger eigenvalues).

Even for small eigenvalues, for example for $\lambda_1 = -\epsilon$, the approximation (12), thus also equation (14), has a limited range of validity. The underlying Padé approximation (19) is valid for small enough s, while for large enough s the exact solution is τ_{tail}. The small s behaviour is captured by the first and second moments, and the large s behaviour, in other terms the tail τ_{tail}, determines the growth of high order moments. This double approximation based on high and low moments may explain why the approximation is successful on diverse data sets (see Fig. 4a). However it is likely to fail precisely when the intermediate moments (third, fourth, etc.), not covered by the approximation, dominate the behaviour of the moment generating function $\rho(s)$. This occurs for instance in the case of a power law of exponent > 3 with sharp cut-off at a large time: while the first and second moment remain bounded and the tail is absent, the intermediate moments grow with-
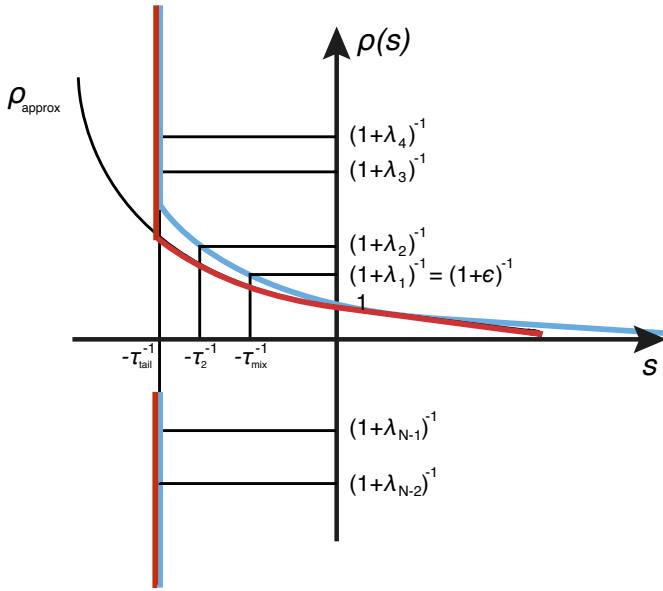
Figure 5: **The exact and approximate characteristic times of a waiting time probability distribution.** Computation of the characteristic time of the distribution $\rho(s) \propto \rho_0(t)e^{-t/\tau_{\text{tail}}}$ (blue curve) with a fat-tailed ρ_0 and a non-analytic point at $s = -1/\tau_{\text{tail}}$. The exact characteristic decay times τ_k are found from the eigenvalues and the blue curve, following equation (18). In this example only three modes, including the stationary, are influenced by the structure of the network, while the other modes collapse to a single decay time τ_{tail}. The dynamics can thus be approximated by three modes, typically describing aggregate probability flows between three network communities. The approximate solution equation (12) follows the red curve, constructed from the Padé approximant ρ_{approx}, see equation (19). As eigenvalues λ_k change from 0 to -2, the values $1/(1 + \lambda_k)$ can take any real value larger than 1 or less than -1.

out bound with the cut-off time. Such distributions are rarely used as models for real-life data however. On the other hand, numerical experiments showed excellent accuracy for a wide range of power-laws with soft cut-off.

Beyond random walks. Random walks are many times formally identical to the asymptotic behaviour of non-linear dynamical models, and also serve as a prototype for a wider class of dynamics that includes epidemic spreading. For example, power networks may be modelled as a network of second-order Kuramoto oscillators[33] whose state in every node is a phase (an angle) θ_i influenced by its neighbours

$$I\ddot{\theta}_i + \alpha\dot{\theta}_i = \omega + \sum_j L_{ij} \sin(\theta_j - \theta_i), \qquad (21)$$

where I is the inertia, α the friction, and ω describes the natural frequency of the oscillation. In this construction, we assume identical parameters for each node. After the change of variables $\phi_i = \theta_i - \omega t/\alpha$, the linearisation around the synchronised equilibrium ($\phi_i = \phi_j$ is constant for all times) is written

$$I\Delta\ddot{\phi}_i + \alpha\Delta\dot{\phi}_i = \sum_j L_{ij}\Delta\phi_j, \qquad (22)$$

which is formally identical to a consensus equation structured by the Laplacian L^{33}. Hence it supports the same formalism as random walks, after the operator D is changed accordingly. The linear dynamics associated to the asymptotic behaviour is often a crucial first step in understanding the non-linear dynamics of the network at some time scale.

Another class of dynamical systems where our results apply are linear dynamics given by equation (1), but where the interaction matrix L does not have a Laplacian structure, and where the dominant eigenmode is not necessarily stationary. This is the case, for instance, for epidemic processes where infected individuals diffuse on a meta-population

network, where nodes are large populations (cities, etc.), and the individuals may either reproduce (by contamination of a healthy individual) or die (or recover). One classic model for this process is a multi-type branching process. This is akin to a random walk whose number of walkers is not preserved in time, as random events are not limited to hops but also death or split. This kind of model also emerges as the linearisation of classic compartmental epidemic models such as SI or SIR[41]. The waiting time between two consecutive events may also be described by an arbitrary distribution $\rho(\Delta t)$. The expected number of walkers in time may again be described by an equation $D\mathbf{x} = L\mathbf{x}$, where D depends on ρ in the same way as before, but now L is no more a Laplacian, and in particular may have negative and positive eigenvalues. If the epidemics is supercritical, then the system is unstable. This implies the existence of unstable eigenmodes ($\lambda_k > 0$) where the number of walkers grows exponentially as e^{t/τ_k}, with τ_k now approximated by $\mu(-\lambda_k^{-1} - \beta)$. In this case, the fat tail of the waiting time distribution does not play any role for the unstable eigenmodes. From this formula, we see that while the decay of stable eigenmodes tends to slow down due to burstiness and possibly fat tail, the unstable eigenmodes are boosted by burstiness. As a consequence, it leads in all eigenmodes to a long-term increase of the infected population due to the temporal heterogeneity.

Numerical analysis. We use six datasets of empirical networks: face-to-face interactions between visitors in a museum (SPM), between conference attendees (SPC)[13], and between hospital staff (SPH)[16]; email communication within a university (EMA)[10]; sexual contacts between sex-workers and their clients (SEX)[14]; and online communication between members of a dating site (POK)[12] (See Table 2).

	SPM	SPC	SPH	EMA	SEX	POK
N	72	113	75	3,186	11,416	28,295
L	6,980	20,818	32,424	309,120	33,645	528,869
δt	min	min	min	hour	day	hour
T (days)	1	2.5	4	82	1,000	512
μ	15.4	80.5	288	61.5	92.9	1,205
σ	30.8	143	502	119	121	1,634
ϵ	0.505	0.709	0.66	0.102	0.012	0.042
τ_{tail}	61	313	664	207	98	1,755

Table 2: **Summary statistics of the empirical networks.** For the largest connected component of the empirical networks used in this study we display the number of vertices (N); number of links (L); temporal resolution used in the simulations (δt); total duration of the network data (T); average (μ) and standard deviation (σ) of waiting times for the simulated random walk, in same unit as δt; and the spectral gap of the original unweighted network (ϵ).

The synthetic networks used in Fig. 2 are formed by 1000 nodes and 10000 links equally divided into two groups, initially disconnected (rewiring probability equal to zero). Within each group, pairs of nodes are formed between uniformly chosen nodes. A fraction of links are rewired to weaken communities. Rewiring consists of uniformly choosing two pairs of nodes and swapping the pairs. Rewiring probability equal to one removes any significant community structure of the network. Communities refer to groups of nodes with more connections between themselves than with nodes at other groups. The spectral gap is calculated for each network configuration. The exemplified waiting time distribution are $\rho(\Delta t) = \exp(-\Delta t/\mu)/\mu$ with $\mu = 1$ (exponential); $= 1.0$ (discrete); $\propto \Delta t^{k-1}\exp(-k\Delta t/\mu)$ with $k = 3$ and $\mu = 3$ (Erlang law); $\propto \Delta t^\gamma$ with $\gamma = 2$ and cutoff at $T = 1000$ (power-law with sharp cut-off); $\propto \Delta t^\gamma \exp(-\Delta t/\tau_{\text{tail}})$ with $\gamma = 3$ and $\tau_{\text{tail}} = 20$ (power-law with soft cut-off).

The synthetic temporal networks used in Figs 3 and 4 are constructed by using the unweighted version of the empirical networks as underlying fixed structures. The randomised networks used in Fig. 3 are obtained by uniformly selecting two pairs of nodes of the original network and then swapping the respective contacts. In the dynamic network, a node and its links alternate between inactive and active states. In order to fit this synthetic temporal network to our theoretical assumptions, we assume that all links of a node are activated

simultaneously, the system is time-invariant (no daily or weekly cycles), exhibits no temporal correlations between successive jumps and node activation times are sampled from the same distribution of inter-activation times (node homogeneity). The active state of a node and its links lasts for one time step δt and then the node and links return to the inactive state. The distribution of inter-activation times (a.k.a. inter-contact times) corresponding to the *original times* is obtained by pooling the times between two subsequent activations of the same node, in other words two consecutive contacts established between the node and its neighbourhood, as observed in the empirical networks[20,43]. The *randomised times* in Fig. 3 correspond to activation times sampled from exponential distributions with the same mean as the corresponding empirical cases. To obtain the waiting time estimates presented in Fig. 3, we simulate a random walk in these synthetic networks. A random walker starts in a randomly chosen node. As time goes by, a walker remains in the node until its new activation and then hops to a uniformly chosen neighbour. The waiting time is thus the time elapsed between the arrival and departure of the walker in a node. We simulate a single random walker and let it hop 300,000 times, starting at 10 uniformly chosen nodes, hence we collect a total of 3 million points for the statistics of waiting times. Besides that, for each starting configuration, we discard the initial 5,000 hops to remove the stochastic transient. The waiting time distribution observed in the simulations provides us with estimates of τ_{mix}, μ, σ^2 and τ_{tail}. The mixing time τ_{mix} is estimated from the distribution expressed in the Laplace domain through equation (18) (for $k = 1$, the mixing mode). This equation is exact since the synthetic data has been constructed so as to satisfy the conditions of validity under which it is proved. One could also estimate the mixing time from the convergence of random trajectories starting in a definite node towards stationarity, however this exercise is computationally demanding. The exponential tail decay time τ_{tail} is estimated using least-square fitting by a line in a lin-log plot of the second half of the waiting time distribution data, thus between $T_{\max}/2$ and $T_{\max}$, where $T_{\max}$ is the largest observed waiting time. Indeed an exponential decay translates into a straight line in a lin-log plot, the slope of which determines the decay time. The choice of the interval $[T_{\max}/2, T_{\max}]$ to perform a linear fit is steered by the want for a simple criterion, uniform across datasets. This may be inappropriate for datasets where the observation time forces a cut-off on the datasets before the tail decay has had time to emerge, for instance in SPC, where a more careful definition of tail, for example on a smaller interval, may be more adequate.

Let us interpret some elementary observations that can be made on Fig. 3. From a) to b), τ_{tail} is only approximately preserved because it is recomputed from the simulated waiting time distribution on a different network, thus a slightly different distribution, even though the inter-contact times are sampled from the same distribution in both panels. In c) and d) the inter-contact time distribution is memoryless (exponential), thus identical to the waiting time distribution regardless of the network, and τ_{tail} coincides with μ. From a) to c), the quantity $\mu \epsilon^{-1}$ drops sharply, although ϵ is clearly preserved as the network is unaltered. The reason is that although the mean inter-contact time is preserved by construction, the mean waiting time also depends on the variance of the inter-contact times, in virtue of the bus paradox, and this variance is clearly strongly affected by the time randomisation. The drop in $\mu \epsilon^{-1}$ from b) to d) is associated to the same mechanism.

To identify the community structure of the empirical networks in Fig. 4, we use the partition stability method[55] using a freely available implementation[63]. The method uses a simple diffusion process such that different communities are detected at different time scales according to the potential of the communities to trap the diffusion at the given time scale. Optimal communities have been derived for values of the resolution parameter decreasing from $10^2, 10^{1.95}, 10^{1.9}, 10^{1.85}, \ldots$, with increasingly fine partitions. After identifying the relevant communities, we discard the inter-community links and calculate the spectral gap of each individual community of at least 10 nodes.

Acknowledgements

Author Contributions

JCD, LECR, and RL conceived the project and wrote the manuscript. JCD derived the analytical results. LECR performed the numerical simulations and analysed the data.

Additional information

References

1. Bansal, S., Read, J., Pourbohloul, B. & Meyers, L. A. The dynamic nature of contact networks in infectious disease epidemiology. *J. Biol. Dynam.* **4**, 478–489 (2010).

2. Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks.* Cambridge University Press, Cambridge (2012).

3. Moody, J. *The Oxford Handbook of Analytical Sociology.* chapter Network Dynamics., pages 447–74. Oxford University Press, Oxford (2009).

4. Newman, M.E.J. *Networks: An Introduction.* Oxford University Press, Oxford (2010).

5. Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nature Phys.* **8**, 32–39 (2012).

6. Chung, F.R.K. *Spectral Graph Theory.* American Mathematical Society (1996).

7. Lovász, L. Random walks on graphs: A survey. *Boy. Soc. Math. Stud.* **2**, 1–46 (1993).

8. Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012).

9. Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).

10. Eckmann, J.-P., Moses, E. & Sergi, D. Entropy of dialogues creates coherent structures in e-mail traffic. *Proc. Nat. Acad. Sci.* **101**, 14333–14337 (2004).

11. Haerter, J.O., Jamtveit, B. & Mathiesen, J. Communication dynamics in finite capacity social networks. *Phys. Rev. Lett.* **109**, 168701 (2012).

12. Holme, P., Edling, C.R. & Liljeros, F. Structure and time-evolution of an Internet dating community. *Soc. Net.* **26**, 2, 155–174 (2004).

13. Isella, L. et al. What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theo. Biol.* **271**, 1, 166–180 (2011).

14. Rocha, L.E.C., Liljeros, F. & Holme, P. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proc. Nat. Acad. Sci.* **107**, 13, 5706–5711 (2010).

15. Starnini, M., Baronchelli, A., Barrat, A. & Pastor-Satorras, R. Random walks on temporal networks. *Phys. Rev. E* **85**, 5, 056115 (2012).

16. Vanhems, P. et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLOS ONE* **8**, 9, e73970 (2013).

17. Rosvall, M., Esquivel, A.V., Lancichinetti, A., West, J.D. & Lambiotte, R. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Comm.* **5**, 4630 (2014).

18. Scholtes, I. et al. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nature Comm.* **5**, 5024 (2014).

19. Holme, P. & Liljeros, F. Birth and death of links control disease spreading in empirical contact networks. *Sci. Rep.* **4**, 4999 (2014).

20. Rocha, L.E.C. & Blondel, V. D. Bursts of vertex activation and epidemics in evolving networks. *PLoS Comput. Biol.* **9**, 3, e1002974 (2013).

21. Horváth, D.X. & Kertész, J. Spreading dynamics on net-

works: the role of burstiness, topology and non-stationarity. *New J. Phys.* **16**, 7, 073037 (2014).

22. Malmgren, R.D., Stouffer, D.B., Motter, A.E. & Amaral, L.A.N. A Poissonian explanation for heavy tails in e-mail communication. *Proc. Nat. Acad. Sci.* **105**, 47, 18153–18158 (2008).

23. Perra, N. et al. Random walks and search in time-varying networks. *Phys. Rev. Lett.* **109**, 238701 (2012).

24. Klafter, J. & Sokolov, I.M. *First Steps in Random Walks: From Tools to Applications.* Oxford University Press, Oxford (2011).

25. Iribarren, J.L. & Moro, E. Impact of human activity patterns on the dynamics of information diffusion. *Phys Rev. Lett.* **103**, 3, 038702 (2009).

26. Jo, H.-H., Perotti, J.I., Kaski, K. & Kertész, J. Analytically solvable model of spreading dynamics with non-Poissonian. *Phys. Rev. X* **4**, 011041 (2014).

27. Min, B., Goh, K.-I. & Vazquez, A. Spreading dynamics following bursty human activity patterns. *Phys. Rev. E* **83**, 3, 036102 (2011).

28. Vazquez, A., Rácz, B., Lukács, A. & Barabási, A.-L. Impact of non-Poissonian activity patterns on spreading processes. *Phys. Rev. Lett.* **98**, 15, 158702 (2007).

29. Aström, K.J. & Murray, R.M. *Feedback Systems: An Introduction for Scientists and Engineers.* Princeton University Press, Princeton (2010).

30. Fax, J.A. & Murray, R.M. Information flow and cooperative control of vehicle formations. *Autom. Cont.* **49**, 9, 1465–1476 (2004).

31. Blondel, V.D., Hendrickx, J.M., Olshevsky, A. & J.N., Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. *Proc. 44th IEEE Conf. Decision Control* , 2996–3000 (2005).

32. Jadbabaie, A., Lin, J. & Morse, A.S. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *Autom. Control, IEEE Trans.* **48**, 6, 988–1001 (2003).

33. Dörfler, F. & Bullo, F. Synchronization and transient stability in power networks and nonuniform Kuramoto oscillators. *SIAM J. Cont. Optim.* **50**, 3, 1616–1642 (2012).

34. Strogatz, S.H. From Kuramoto to Crawford: Exploring the onset of synchronization of in populations of coupled oscillators. *Phys. D* **143**, 1-4, 1–20 (2000).

35. Yin, G.G. & Zhang, Q. *Continuous-time Markov Chains and Applications: A Two-time-scale Approach.* volume 37. Springer (2012).

36. Levin, D.A., Peres, Y. & Wilmer, E.L. *Markov Chains and Mixing Times.* American Mathematical Society (2008).

37. Hoffmann, T., Porter, M.A. & Lambiotte, R. Generalized master equations for non-Poisson dynamics on networks. *Phys. Rev. E* **86**, 4, 046102 (2012).

38. Montroll, E.W. & Weiss, G.H. Random walks on lattices ii. *J. Math. Phys.* **6**, 2, 167–181 (1965).

39. Kolmogoroff, A. Uber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* **104**, 1, 415–458 (1931).

40. Dyke, P. *An Introduction to Laplace Transforms and Fourier Series.* Springer (2014).

41. Keeling, M.J. & Ross, J.V. On methods for studying stochastic disease dynamics. *J. R. Soc. Interface* **5**, 19, 171–181 (2008).

42. Kivelä, M. et al. Multiscale analysis of spreading in a large communication network. *J. Stat. Mech.* **2012**, 03, P03005 (2012).

43. Rocha, L.E.C., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol.* **7**, 3, e1001109 (2011).

44. Speidel, L., Lambiotte, R., Aihara, K. & Masuda, N. Steady state and mean recurrence time for random walks on stochastic temporal networks. *Phys. Rev. E* **91**, 1, 012806 (2015).

45. Kleinrock, L. Queueing systems: Theory. I. (1975).

46. Lambiotte, R., Tabourier, L. & Delvenne, J.-C. Burstiness and spreading on temporal networks. *Eur. Phys. J. B* **86**, 7 (2013).

47. Karsai, M. et al. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 2, 025102 (2011).

48. Goh, K.I. & Barabási, A.-L. Burstiness and memory in complex systems. *Europhys. Lett.* **81**, 4 (2008).

49. Cheeger, J. *Problems in Analysis.* chapter A lower bound for the smallest eigenvalue of the Laplacian., pages 195–199. Princeton University Press, Princeton, N.J. (1970).

50. Diaconis, P. & Stroock, D. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1**, 1, 36–61 (1991).

51. Gfeller, D. & De Los Rios, P. Spectral coarse graining and synchronization in oscillator networks. *Phys. Rev. Lett.* **100**, 17, 174104 (2008).

52. Kokotovic, P., Khalil, H.K. & O'reilly, J. *Singular Perturbation Methods in Control: Analysis and Design.* volume 25. Society for Industrial and Applied Mathematics (1987).

53. Simon, H.A. & Ando, A. Aggregation of variables in dynamic systems. *Econometrica* , 111–138 (1961).

54. Simonsen, I. Diffusion and networks: A powerful combination! *Phys. A* **357**, 2, 317–330 (2005).

55. Delvenne, J.-C., Schaub, M.T., Yaliraki, S.N. & Barahona, M. *Dynamics On and Of Complex Networks.* volume 2, chapter The stability of a graph partition: A dynamics-based framework for community detection., pages 221–242. Springer (2013).

56. Shen, H.-W. & Cheng, X.-Q. Spectral methods for the detection of network community structure: A comparative analysis. *J. Stat. Mech.* **2010**, 10, P10020 (2010).

57. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **17**, 4, 395–416 (2007).

58. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).

59. Ronhovde, P. & Nussinov, Z. Multiresolution community detection for megascale networks by information-based replica correlations. *Phys. Rev. E* **80**, 1, 016109 (2009).

60. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).

61. Karrer, B., Levina, E. & Newman, M.E.J. Robustness of community structure in networks. *Phys. Rev. E* **77**, 4, 046119 (2008).

62. Delmotte, A., Tate, E.W., Yaliraki, S.N. & Barahona, M. Protein multi-scale organization through graph partitioning and robustness analysis: Application to the myosin light chain interaction. *Phys. Biol.* **8**, 5, 055010 (2011).

63. Delmotte, A., Schaub, M.T., Yaliraki, S.N. & Barahona, M., Community Detection using the stability of a graph partition. https://github.com/michaelschaub/PartitionStability (2012).

64. H. Miyazako, Y. Hori, S. Hara Turing instability in reaction-diffusion systems with a single diffuser: Characterization based on root locus. in: Dec. Cont. (CDC), IEEE 52nd Ann. Conf. 2671–2676 (2013).

65. S. Brin, L. Page The anatomy of a large-scale hypertextual Web search engine. Comp. Net. ISDN Syst. 30 (1) 107–117 (1998).

66. M. Rosvall, C. T. Bergstrom Maps of random walks on complex networks reveal community structure., Proc. Nat. Acad. Sci. 105 (4) 1118–1123 (2008).

67. H. W. Hethcote The mathematics of infectious diseases.

SIAM Rev. 42 (4) 599–653 (2000).

68. P. J. Menck, J. Heitzig, N. Marwan, J. Kurths How basin stability complements the linear-stability paradigm. Nature Phys. 9 (2) 89–92 (2013).

SUPPLEMENTARY NOTE 1

We extend the discussion on the main text on power-law distribution of waiting times and show results for various values of parameters, including the transition thresholds between dominating factors. Power-law functions represent a particularly important family of probability distributions. Although many times power-laws are not the optimal statistical probability distribution to model waiting times of human interactions, they properly capture both aspects of realistic distributions, i.e. fat tails and burstiness, in a single functional form, and is in contrast to standard models of homogeneous (e.g. exponential) and regular times.
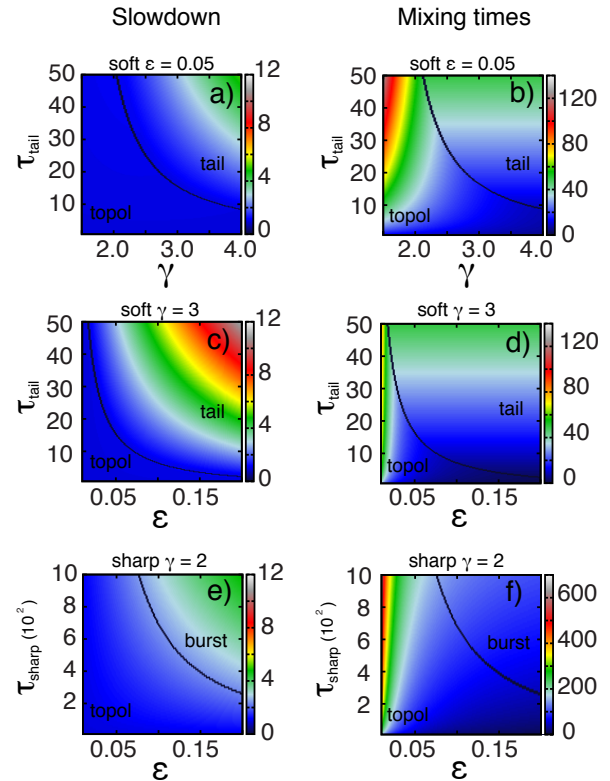

Figure 6: **Slowdown factor and mixing times for power-law distributions of waiting times.** Panels (a,b) correspond to strong clustering (fixed $\epsilon = 0.05$) and (c,d) to fixed $\gamma = 3$ for soft cutoff; and (e,f) correspond to fixed $\gamma = 2$ for sharp cutoff. Black curves divide the phases where either structure (topology), or burstiness, or tail dominates. The slowdown factor is the ratio between the mixing time of the power-law distribution of waiting times and of the exponential distribution (Poisson process). The value of the mixing time is taken as $\max(\mu(\epsilon^{-1} + \beta), \tau_{\text{tail}})$, always within 5% of the exact value as computed with equation (18) of the main text.

As presented in the main text, in the extreme case of waiting times following a pure power-law, i.e. $\tau_{\text{tail}} = \infty$, infinite mixing time occurs irrespective of the network structure, as convergence to stationary state takes place at subexponential rate. This can be shown directly as after any time t, there will be a probability $1 - \int_0^t \rho(r)dr$, polynomially decreasing, that the random walker has not moved at all from its initial node, making exponential convergence to stationarity impossible. If a soft (exponential) cutoff is added so that $\rho(t) \propto (t+1)^{-\gamma} \exp(-t/\tau_{\text{tail}})$, the slowdown is either driven by the tail or by the structure (Figure 6b,d). Above a certain ϵ^*, the tail completely determines the mixing time, which is not surprising considering that large ϵ corresponds to the absence of bottlenecks or strong clustering. Below this threshold, for instance at $\epsilon = 0.05$ (Figure 6a,b), there is a strong dependence on the structure for a range of exponents γ and τ_{tail}. In this scenario, burstiness never dominates, and the power-law exponent γ does not play a direct role other than influencing μ. Nevertheless, if the distribution of waiting times is modelled by a power-law with sharp cutoff,

e.g. $\rho(t) \propto (t+1)^{-\gamma}$ for $0 \leq t \leq \tau_{\mathrm{sharp}}$, and $\rho(t) = 0$ for $t > \tau_{\mathrm{sharp}}$, there is no tail, and only burstiness and structure compete to regulate the slowdown (Figure 6e,f).

These simple theoretical models allow to validate the accuracy of the approximation given by equation (12) in the main text for the mixing mode, $\tau_{\mathrm{mix}} \approx \max(\mu(\epsilon^{-1} + \beta), \tau_{\mathrm{tail}})$, to estimate the mixing time, as compared to the exact formula, see equation (18) in Methods B (error $< 5\%$ on the examples of Figure 6). They also allow to illustrate the fact that burstiness (dominating sharp cut-off power-laws) and fat tails (dominating soft cut-off power-laws) are independent features of a distribution of waiting times, as already exemplified in the main text.

SUPPLEMENTARY NOTE 2

An empirical distribution of waiting times necessarily vanishes after some observation time T. Assume that the actual distribution is $\rho(\Delta t) = \alpha \rho_0(\Delta t) e^{-\Delta t / \tau_{\mathrm{tail}}}$, where $\rho_0(\Delta t)$ is a fat tailed distribution and α a normalising constant. Assimilate the empirically observed distribution ρ_T to the truncation $\rho(\Delta t) / \int_0^{T_{\mathrm{obs}}} \rho$ until $\Delta t = T$ and zero for $\Delta t > T$. Then as T grows to infinity, the moments of empirical and actual distribution will coincide. As for the empirical Laplace transform $\rho_T(s)$, it is defined and analytic on the whole complex plane. We know that the Laplace transform of the actual distribution $\rho(\Delta t)$ cannot be extended to the left of $s = -1/\tau_{\mathrm{tail}}$. What happens to $\rho_T(s)$ around those values? Choosing $s = -1/\tau_{\mathrm{tail}} - \Delta s$, we write $\rho_T(s) = \int \rho_T(\Delta t) e^{-s\Delta t} d\Delta t$. If assume for simplicity $\rho_0(\Delta t)$ to be decreasing, then a simple calculation yields $\rho_T(s) \geq \alpha \rho_0(T)(e^{\Delta s T} - 1)/\Delta s$. Therefore $\rho_T(s)$, for s even slightly below $-1/\tau_{\mathrm{tail}}$, increases exponentially with T, thus is practically vertical for a high enough value of T. Therefore whenever a mode has characteristic time $\tau_k = \tau_{\mathrm{tail}}$, the computation of τ_k from the empirical distribution and equation (17-18) in Methods B, $\rho_T(-1/\tau_k) = (1 + \lambda_k)^{-1}$, is expected to be close to $\tau_k \approx \tau_{\mathrm{tail}}$ as well. We deduce that provided a high enough observation time, the modelling of the distribution by a truncated distribution over the finite interval or by the extrapolation to a form $\rho_0(\Delta t) e^{-\Delta t / \tau_{\mathrm{tail}}}$ is irrelevant, as both will lead to the same conclusions.

This fact leads to an interesting self-consistency condition as if the observation time is sufficiently long, we find that both the truncated model ρ_T (via the equation $\rho_T(-1/\tau_{\mathrm{mix}}) = (1 - \epsilon)^{-1}$) or the full-timeline model ρ (leading to $\tau_{\mathrm{mix}} = \tau_{\mathrm{tail}}$ for fat-tail dominated networks) conclude to the same mixing time. If in addition, the approximation of equation (12) of the main text is valid for the mixing mode ($k = 1$), we find the equation

$$\tau_{\mathrm{mix}} \approx \mu(\epsilon^{-1} + \beta) \approx \tau_{\mathrm{tail}}, \qquad (23)$$

indeed approximately verified by the face-to-face networks dominated by fat tail effect (see Fig. 3), especially SPM and SPH. A strong discrepancy between the sharp cut-off and full timeline model may reveal an excessively short observation time, insufficient for a reliable estimation of the mean, variance and/or tail of the distribution of waiting times, as is the case for SPC.

SUPPLEMENTARY NOTE 3

The mathematical framework proposed in the main text adresses the combined effect of the network topology and the distribution of waiting times. Temporal networks obtained from real data however contain a number of temporal correlations and patterns, such as non-stationarity[19–21] and causality[18], or structure-time correlations[42,47]. A careful analysis of the various patterns and correlations present in real data is out of the scope of our analysis. In order to contrast the limitations of our theory to the dynamics in real settings, we test our

estimations by simulating a random walk dynamics directly on the empirical network such that the effect of all correlations and patterns of both time and structure are captured.

In the empirical temporal network each link has a time-stamp to indicate exactly the time it is active (being inactive at other times). The link to empirical temporal networks is done by performing a random walk on the time-stamped empirical data. We start a large number of walkers on each node at $t = 0$. As time goes by, a walker remains in the node until one of its links becomes active and then jumps to the corresponding neighbour. The potential paths for the random walk are entirely defined by the empirical sequence of contacts (capturing temporal and temporal/structure correlations), meaning that not only the

	δt	μ	σ	τ_{tail}	ϵ	β	τ_{mix}	M
SPM	min	3.3	15.7	48.8	0.142	10.8	59.0	Tail
SPC	min	16.2	101.2	182.5	0.073	19.0	529.8	Burstiness
EMA	hour	14.5	59.4	177.4	0.074	7.9	310.4	Structure
SEX	week	18.0	32.2	20.2	0.012	1.1	1519.8	Structure

Table 3: **Summary statistics of the empirical networks.** Temporal resolution (δt); mean (μ) and standard deviation (σ) of the waiting time distribution; τ_{tail} is the least-square estimate of the tail characteristic time of the cumulative distribution of waiting times; ϵ is the spectral gap for the aggregated weighted network; τ_{mix} is formally computed as $\max(\mu\epsilon^{-1}, \mu\beta, \tau_{\mathrm{tail}})$ in analogy with the main text's case, even though this is not necessarily representative of the actual mixing time in this situation; and M refers to the regulating mechanism that determines τ_{mix}. We use the largest connected component.

inter-contact time distribution is regulating the diffusion. This is in contrast to the experiments presented in the main text, where the dynamic networks had the empirical structure (unweighted) fixed and the activation times of nodes were sampled from the empirical inter-contact time distribution (this process generates uncorrelated activation times following the empirical inter-contact time distribution). The spectral gap is calculated using the weighted version of the network, therefore inhomogeneities in the weight-topology are captured. The distribution of waiting times of the walkers provides us with empirical measures of μ, σ^2 and τ_{tail}. We apply this process to four of the datasets studied in the main text: face-to-face interactions between visitors in a museum (SPM) and between conference attendees (SPC)[13], to email communication within a university (EMA)[10] and to sexual contacts between sex-workers and -buyers (SEX)[14] (see Table 3). We underline that the two experiments, from this section and from Fig. 3, use different methodologies, as the latter removed correlations or non-stationarity, but also weights on the network (allowing to perform structure randomisation in Fig. 3b,d – main text – conveniently). Therefore they are not supposed to deliver the same temporal or structural characteristics, except in order of magnitude. Moreover the mixing time in the diffusion process involving all correlations is not necessarily given by the formula $\tau_{\mathrm{mix}} \approx \max(\mu\epsilon^{-1}, \mu\beta, \tau_{\mathrm{tail}})$, which should be updated to account for event-event correlations, daily or weekly periodicity, etc. Nevertheless among those three mechanisms included in the formula, we notice a domination of waiting time effects, tail or burstiness in the SPC and SPM datasets. On the other hand, structural bottlenecks seem to be strong enough to dominate waiting times effects for EMA and SEX datasets, consistently with the experiment results in Fig. 3 in the main text. In future work, it would be interesting to check if supplementary correlations may reverse the respective strength of structure and waiting time effects.