THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

Hierarchy of temporal responses of multivariate self-excited epidemic processes

Alexander Saichev[1,2], Thomas Maillart[3], and Didier Sornette[1,a]

[1] Department of Management, Technology and Economics, ETH Zurich, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
[2] Mathematical Department, Nizhny Novgorod State University, Gagarin prosp. 23, 603950 Nizhny Novgorod, Russia
[3] Center for Law & Economics, ETH Zurich, Haldeneggsteig 4, 8092 Zurich, Switzerland

Abstract. Many natural and social systems are characterized by bursty dynamics, for which past events trigger future activity. These systems can be modelled by so-called self-excited Hawkes conditional Poisson processes. It is generally assumed that all events have similar triggering abilities. However, some systems exhibit heterogeneity and clusters with possibly different intra- and inter-triggering, which can be accounted for by generalization into the "multivariate" self-excited Hawkes conditional Poisson processes. We develop the general formalism of the multivariate moment generating function for the cumulative number of first-generation and of all generation events triggered by a given mother event (the "shock") as a function of the current time t. This corresponds to studying the response function of the process. A variety of different systems have been analyzed. In particular, for systems in which triggering between events of different types proceeds through a one-dimension directed or symmetric chain of influence in type space, we report a novel hierarchy of intermediate asymptotic power law decays $\sim 1/t^{1-(m+1)\theta}$ of the rate of triggered events as a function of the distance m of the events to the initial shock in the type space, where $0 < \theta < 1$ for the relevant long-memory processes characterizing many natural and social systems. The richness of the generated time dynamics comes from the cascades of intermediate events of possibly different kinds, unfolding via random changes of types genealogy.

1 Introduction

We study a class of point processes that was introduced by Hawkes in 1971 [1–4]. It is much richer and relevant to most natural and social systems than standard point processes [5–10], because it describes "self-excited" processes. This term means that the past events have the ability to trigger future events, i.e., the intensity $\lambda(t|H_t)$ (conditional mean instantaneous rate) of the point process is a function of past events, being therefore non-markovian. Many works have been performed to characterize the statistical and dynamical properties of this class of models, with applications ranging from geophysical [11–16], medical [17] to financial systems, with applications to Value-at-Risk modeling [18], high-frequency price processes [19], portfolio credit risks [20], cascades of corporate defaults [21], financial contagion [22], and yield curve dynamics [23].

However, in reality, in many systems, events come in different types with possibly different properties, while keeping a degree of mutual inter-excitations. Indeed, types of events and their triggering effects vary with the conditions under which they have initiated: earthquakes and aftershocks are partitioned in different tectonic

regions [24], financial volatility bursts occur at given time periods on some particular classes of assets [25], with loose or no influence on others, and information and memes might percolate in social media and influence variously online communities [26,27]. However, the widely studied mono-variate approach does not capture the hierarchy of different interactions between plates, assets and social communities, because it assumes that all events are of the same type.

To motivate the relevance of the multivariate version of the self-excited Hawkes conditional Poisson processes, we present a concrete example, namely the dynamics of open source software developments. Open source software (OSS) is a paramount and reproducible example of collective action [28] with lots of developers contributing source code over generally very long time periods. As in many – if not all – human systems, OSS exhibits bursty dynamics of contributions at various coarse-grained levels. Figure 1 shows some example of the dynamics of contributions at the developer level over approximately two years, which reveals bursty as well as some partial synchronization of actions between some developers.

It is essential to note that the community of each OSS project is very heterogeneous. There exists a hierarchy of clusters in such social networks [29] so that (i) developers
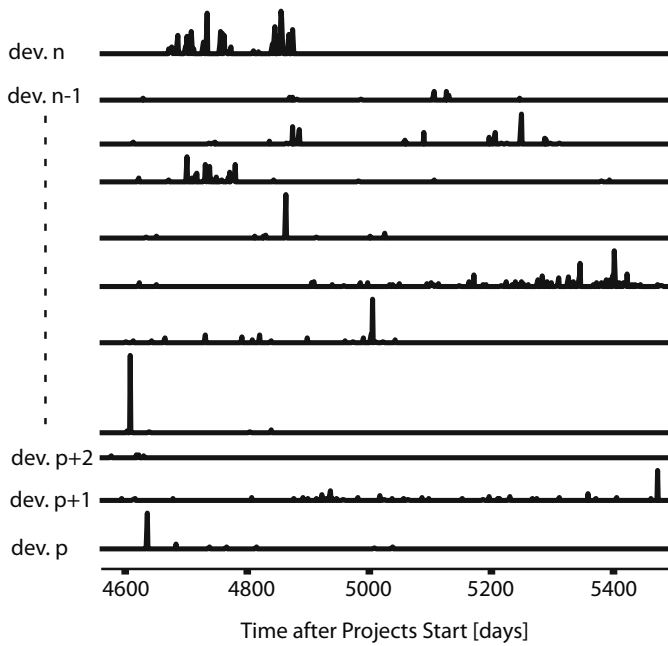
[a] e-mail: dsornette@ethz.ch

Fig. 1. Typical time series of the activities of developers quantified by the (normalized) number of commits per day taken for the Eclipse Project. Each line shows the activity of one developer over approximately two years. Simple visual inspection shows the presence of coherent dynamics, spiky dynamics and clustering of activities between developers.

Fig. 2. (Color online) Scatter plot of the activities of developers over time for one single file in the Mozilla Project. This example is similar to thousands of files found in many open source projects. On the ordinate axis, developers are indexed by their level of seniority (developer #1 is the most senior). The area of each circle shows the number of times this file has been modified by a single developer over a given period (here time is binned by periods of 14 days). The green squares show the commits of the most active developer for this file. On the one hand, commit activity exhibits (i) long memory processes; (ii) triggering within developer's activity; (iii) between developers; and (iv) between groups of developers (generally clustered by seniority). On the other hand, many developers remain marginally or completely absent from the development of a specific part of the source code.

specialize horizontally and vertically [30] according to their respective motivations [31]; and (ii) the modular structure of source code [32,33] further constraints and propagates interactions depending on the centrality of a component [34,35].

Therefore, observed social networks in OSS are more the result of various levels of interaction between developers constrained by their motivations, programming skills and expertise, and the structure of the code itself. For these reasons, two developers in a same project might never interact. A contrario, one developer (resp. a group of developers) might change the code in a way that obliges all other developers to make changes to their own code. Moreover, a developer might only make changes following others changes (e.g. debugger).

Figure 2 illustrates the dynamics of interaction between developers for a file in Mozilla, which has been changed regularly over time. The scatter plot shows the number of commits per 14 days (disk area) by developer (ordered by seniority on the ordinate axis) and as a function of time (on the abscissa). Figure 2 clearly exhibits clustering in these different dimensions: commits are bursty over time as a result of long memory processes [36–38], and developers coordinate to change source code files. Moreover, it appears that developers cluster by seniority. Similarly, Figure 3 presents the dynamics of changes for one developer in multiple files. While the patterns are slightly different in details, clustering and long memory processes occur as well.
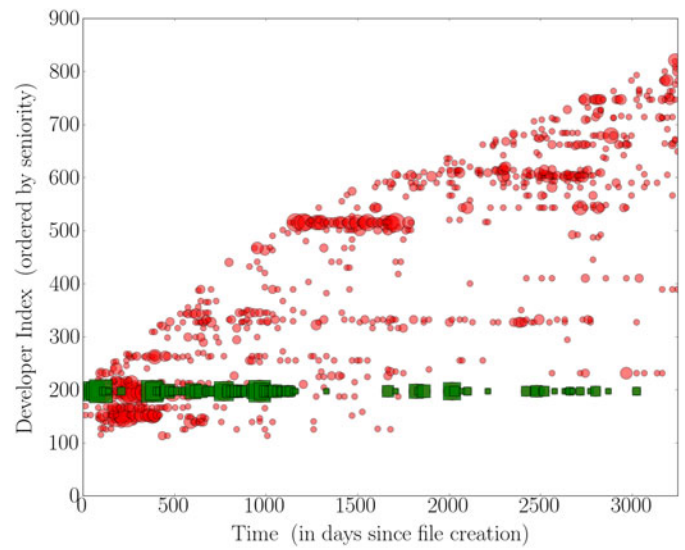
Figures 1–3 illustrate that the dynamics of open source software development occurs by interactions between developers at many different levels with various interaction strengths that result from different levels of involvement, programming skills, areas of expertise and interest, and the degree of advancement of the project. Also, as a result of the structure of the system itself, building blocks (modules) are either highly central with critical implications when changed, or peripheral with less implications. These categories form clusters of developers, which interact with various triggering characteristics. While triggering within one cluster might be high due to homogeneity (e.g. people talk the same language, know each other, trust between actors has developed), triggering between clusters might be smaller or higher, mono-directional or bi-directional, etc. These observations are summarized in the conceptual cartoon shown in Figure 4 in which the tent-like patterns represent the same-developer endogenous triggering while the arrows indicate the possibility for mutual triggering between different developers. This is precisely the co-existence of the triggering of the coding activities at the intra-developer and inter-developer levels that the present multivariate Hawkes process aims at modeling. This co-existence of intra- and inter-excitations
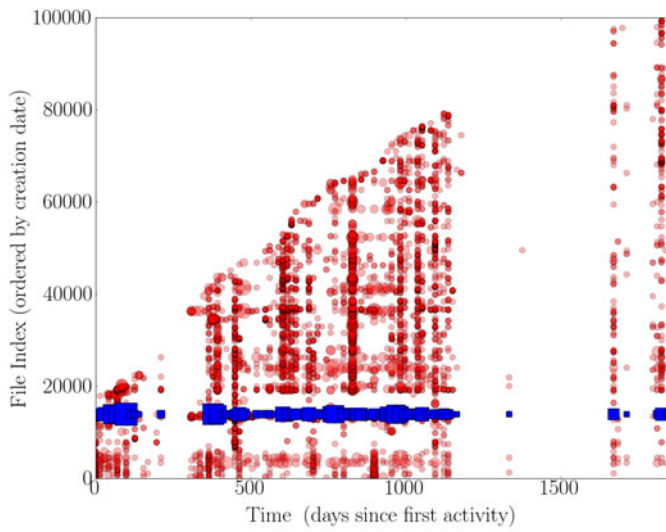
Fig. 3. (Color online) Scatter plot of the activities of files over time for the developer who contributed the most for the file used in Figure 2 (presented as blue squares here). This example is similar to the behaviors observed for hundreds of developers in many open source projects. On the ordinate axis, files are indexed by their creation date (file #1 is the oldest). The area of each circle shows the number of times a file has been modified over a given period (here time is binned by periods of 14 days). Similarly to Figure 2, the activities of developers exhibit clear clustering and memory effects.

Fig. 4. (Color online) Conceptual representation of intra-developer and inter-developer activity triggering. The tent-like patterns represent the same-developer endogenous triggering while the arrows indicate the possibility for mutual triggering between different developers.

illustrated in Figure 4 is generic to the many systems mentioned above and many more.

These observations suggest that multivariate self-excited point processes, which extend the class of monovariate self-excited point processes, provide an adequate class of models to describe the self-excitation (or intra-triggering) as well as the mutual influences or triggering between different types of events that occur in many natural and social systems. Indeed, they capture exogenous shocks, inter- and intra-triggering endogenous shocks with memory, as usually observed. Actually, the generalization to multivariate self-excited point processes was mentioned by Hawkes himself in his first paper [1], but the full relevance of this class of models has only been recently appreciated [22,39]. In the context of open source software, but not only there, multivariate self-excited point processes offer a convenient tool to reconstruct and to predict social and technical interactions between developers over the course of OSS projects.

The organization of the paper proceeds as follows. Section 2 first recalls the definition of the monovariate Hawkes process and then presents the general multivariate self-exciting Hawkes processes. Section 3 develops the formalism of the multivariate moment generating function for the cumulative number of first-generation and of all generation events triggered by a given mother event as a function of the current time t. It also provides the general relations to obtain the mean numbers of events triggered over all generations by a given event as a function of time. Sect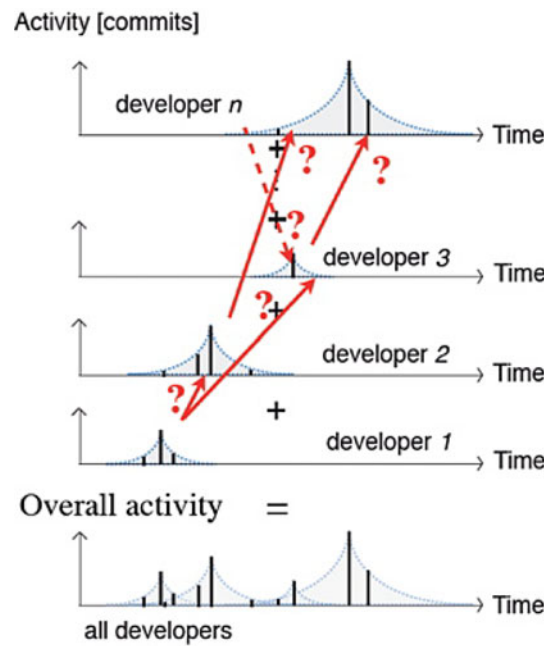ion 4 analyzes a system in which the intra-type triggering processes are all of the same efficiency while all inter-type triggering processes have themselves the same efficiency between themselves but in general weaker than the intra-type triggering processes. We derive the time-dependence of the rates of events triggered from a given shock for distributions of waiting times of first-generation events that have either exponential or power law tails. Section 5 analyzes a system in which triggering between events of different types proceeds through a one-dimension directed chain of influence in type space. We uncover a novel hierarchy of intermediate asymptotic power law decays of the rate of triggered events as a function of the distance of the events to the initial shock in the space of types. Section 6 generalizes the results of Section 5 by studying a system in which triggering between events of different types proceeds through a one-dimension symmetric chain of influences in type space. Section 7 concludes and two Appendices give proofs and details of the key results of the paper.

2 Definitions and notations for the multivariate Hawkes processes

2.1 Monovariate Hawkes processes

Self-excited conditional Poisson processes generalize the cluster models by allowing each event, including cluster members, i.e., aftershocks, to trigger their own events

according to some memory kernel $h(t - t_i)$:

$$\lambda(t|H_t, \Theta) = \lambda_c(t) + \sum_{i|t_i<t} f(t - t_i), \qquad (1)$$

where the history $H_t = \{t_i\}_{1\leq i\leq i_t,\ t_{i_t}\leq t<t_{i_t+1}}$ includes all events that occurred before the present time t and the sum in expression (1) runs over all past triggered events. The set of parameters is denoted by the symbol Θ. The term $\lambda_c(t)$ means that there are some external background sources occurring according to a Poisson process with intensity $\lambda_c(t)$, which may be a function of time, but all other events can be both triggered by previous events and can themselves trigger their offsprings. This gives rise to the existence of many generations of events.

Introducing "marks" or characteristics for each event leads to a first multidimensional extension of the self-excited process (1). The generalization consists in associating with each event some marks (possible multiple traits), drawn from some distribution $p(m)$, usually chosen invariant as a function of time:

$$\lambda(t, M|H_t, \Theta) = p(M)\left(\lambda_c(t) + \sum_{i|t_i<t} f(t - t_i, M_i)\right), \quad (2)$$

where the mark M_i of a given previous event now controls the shape and properties of the triggering kernel describing the future offsprings of that event i. The history now consists in the set of occurrence times of each triggered event and their marks: $H_t = \{t_i, M_i\}_{1\leq i\leq i_t,\ t_{i_t}\leq t<t_{i_t+1}}$. The first factor $p(M)$ in the r.h.s. of expression (2) writes that the marks of triggered events are drawn from the distribution $p(M)$, independently of their generation and waiting times. This is a simplifying specification, which can be relaxed. Inclusion of a spatial kernel to describe how distance impacts triggering efficiency is straightforward.

A particularly well-studied specification of this class of marked self-excited point process is the so-called Epidemic-Type-Aftershock-Sequence (ETAS) model [12,40][1]:

$$\lambda(t, M|H_t, \Theta) = p(M)\left(\lambda_c + \sum_{i|t_i<t} \frac{ke^{a(M_i-M_0)}}{(t - t_i + c)^{1+\theta}}\right), \quad (3)$$

---

[1] This paper offers a continuum-state critical branching process which develops along the time axis. The continuum-state means that earthquake events are infinitesimal and to obtain 'real' earthquakes, each realization is processed through a special filter. This paper can however be considered as an ancestor to Ogata's ETAS model [12], because it has the same branching structure of jumps and power law triggering function. Apart from being used differently (Kagan and Knopoff [40] to generate one complex earthquake and Ogata [12] to generate a complex sequence of aftershocks triggered by a main earthquake), the only real difference between the two models is the regularization of the memory kernel at short times, which has no significant impact.

where $p(m)$ is given by the Gutenberg-Richter law, which describes the probability density function (pdf) of earthquakes of a given energy E, as being a power law, which translates into an exponential function $p(M) \sim e^{-bM}$ of the log-energy scales inscribed in magnitudes M's. The memory kernel is often chosen as the power law (called the Omori law) with exponent $1 + \theta$ and usually $0 \leq \theta < 1$. The time constant c ensures finiteness of the triggering rate immediately following any event. Other forms with shorter memory, as the exponential, are also common. Each event (of magnitude $M \geqslant M_0$) triggers other events with a rate $\sim e^{aM}$, which defines the so-called fertility or productivity law. The lower magnitude cut-off M_0 is such that events with marks smaller than M_0 do not generate offsprings. This is necessary to make the theory convergent and well-defined, otherwise the crowd of small events may actually dominate [41,42]. The constant k controls the overall productivity law and thus the average branching ratio defined by expression (4) below. The set of parameters is $\Theta = \{b, \lambda_c, k, a, M_0, c, \theta\}$.

From a theoretical point of view, the Hawkes models with marks has been studied in essentially two directions: (i) statistical estimations of its parameters with corresponding residual analysis as goodness of fits [43–52]; (ii) statistical properties of its space-time dynamics [11,16,53–62].

The advantage of the self-excited conditional Hawkes process is to provide a very parsimonious description of the complex spatio-temporal organization of systems characterized by self-excited bursts of events, without the need to invoke ingredients other than the generally well-documented stylized facts on the distribution of event sizes, the temporal "Omori law" for the waiting time before excitation of a new event and the productivity law controlling the number of triggered events per initiator.

Self-excited models of point processes with additive structure of their intensity on past events as in (2) and (3) [4] make them part of the general family of branching processes [63]. The crucial parameter is then the average branching ratio n, defined as the mean number of events of first generation triggered per event. Using the notation of expression (3), the average branching ratio is given by

$$n = \frac{k}{\theta c^\theta}\frac{b}{b - a}. \qquad (4)$$

Depending on applications, the branching ratio n can vary with time, from location to location and from type to type (as we shall see below for the multivariate generalization). The branching ratio provides a diagnostic of the susceptibility of the system to trigger activity in the presence of some exogenous nucleating events.

Precise analytical results and numerical simulations show the existence of three time-dependent regimes, depending on the "branching ratio" n and on the sign of θ. This classification is valid for the range of parameters $a < b$. When the productivity exponent a is larger than the exponent b of the Gutenberg-Richter law, formula (4) does not make sense anymore, which reflects the existence of an explosive regime associated with stochastic finite-time

singularities [61], a regime that we do not consider further below, but which is relevant to describe the accelerated damage processes leading to global systemic failures in possibly many different types of systems [64].

1. For $n < 1$ (sub-critical regime), the rate of events triggered by a given shock decays according to an effective Omori power law $\sim 1/t^p$, characterized by a crossover from an Omori exponent $p = 1-\theta$ for $t < t^*$ to a larger exponent $p = 1+\theta$ for $t > t^*$ [11], where t^* is a characteristic time $t^* \simeq c/(1-n)^{1/\theta}$, which is controlled by the distance of n to 1.
2. For $n > 1$ and $\theta > 0$ (super-critical regime), one finds a transition from an Omori decay law with exponent $p = 1 - \theta$ at early times since the mainshock to an explosive exponential increase of the activity rate at times $t > t^* \simeq c/(n-1)^{1/\theta}$ [11,65].
3. In the case $\theta < 0$, there is a transition from an Omori law with exponent $1 - |\theta|$ similar to the local law, to an exponential increase at large times, with a crossover time τ different from the characteristic time t^* found in the case $\theta > 0$.

We refer in particular to reference [17] for a short review of the main results concerning the statistical properties of the space-time dynamics of self-excited marked Hawkes conditional Poisson processes.

2.2 Multivariate Hawkes processes

The Multivariate Hawkes Process generalizes expressions (2) and (3) into the following general form for the conditional Poisson intensity for an event of type j among a set of m possible types (see the document [66] for an extensive review):

$$
\lambda_j(t|H_t) =
$$
$$
\lambda_j^0(t) + \sum_{k=1}^m \Lambda_{kj} \int_{(-\infty,t)\times\mathcal{R}} f_{k,j}(t-s)g_k(x)N_k(ds \times dx),
$$
$$(5)$$

where H_t denotes the whole past history up to time t, λ_j^0 is the rate of spontaneous (exogenous) events of type j, i.e., the sources or immigrants of type j, Λ_{kj} is the (k,j)'s element of the matrix of coupling between the different types which quantifies the ability of a type k-event to trigger a type j-event. Specifically, the value of an element Λ_{kj} is just the average number of first-generation events of type j triggered by an event of type k. This generalizes the branching ratio n defined by (4). The memory kernel $f_{k,j}(t-s)$ gives the probability that an event of type k that occurred at time $s < t$ will trigger an event of type j at time t. The function $f_{k,j}(t-s)$ is nothing but the distribution of waiting times $t - s$ between the impulse of event k which impacted the system at some time s and the occurrence of an event of type j at time t. The fertility (or productivity) law $g_k(x)$ of events of type k with mark x quantifies the total average

number of first-generation events of any type triggered by an event of type k. We have used the standard notation $\int_{(-\infty,t)\times\mathcal{R}} f(t,x)N(ds \times dx) := \sum_{i|t_i<t} f(t_i, x_i)$.

The matrix Λ_{kj} embodies both the topology of the network of interactions between different types, and the coupling strength between elements. In particular, Λ_{kj} includes the information contained on the adjacency matrix of the underlying network. Analogous to the condition $n < 1$ (subcritical regime) for the stability and stationarity of the monovariate Hawkes process, the condition for the existence and stationarity of the process defined by (5) is that the spectral radius of the matrix Λ_{kj} be less than 1. Recall that the spectral radius of a matrix is nothing but its largest eigenvalue.

To our knowledge, all existing works on the multivariate Hawkes processes assume that the first moment of the memory kernel $f_{k,j}(t-s)$ exists. In the notations analogous to those of equation (3), if the memory kernels have an Omori like power law tail $\sim 1/t^{1+\theta}$, this first-order moment condition imposes that $\theta > 1$. But, this is not the correct regime of parameters for earthquakes as well as for other social epidemic processes, which have been shown to be characterized by long-memory processes with $0 < \theta < 1$ [67–70]. This regime $0 < \theta < 1$ leading to infinite first-order moments leads to very rich new scaling behaviors in the multivariable case, as we are going to show below. Actually, we will derive the remarkable results that multivariate Hawkes processes can be characterized by a hierarchy of dynamics with different exponents, all related to the fundamental Omori law for first generation waiting times.

3 Temporal multivariate moment generating function (MGF)

While the technology of moment generating functions is well-known for general branching processes [7,63], the contribution of this section is to make explicit how to apply it for the temporal properties of the multivariate Hawkes process.

3.1 Moment generating function for the cumulative number of first-generation events triggered until time t

Among the m types of events, consider the kth type and its first generation offsprings. Let us denote $R_1^{k,1}(t), R_1^{k,2}(t), \ldots, R_1^{k,m}(t)$, the cumulative number of "daughter" events of first generation of type $1, 2, \ldots, m$ generated by this "mother" event of type k from time 0 until time t. With these notations, the moment generating function (MGF) of all events of first generation that are triggered by a mother event of type k until time t reads

$$
A_1^k(\mathbf{y}; t) := \mathrm{E}\left[\prod_{s=1}^m y_s^{R_1^{k,s}(t)}\right],
$$
$$(6)$$

where $\mathrm{E}\left[X\right]$ denotes the average of X over all possible statistical realizations. We have introduced the vector notation $\mathbf{y} \equiv (y_1, y_2, \ldots, y_m)$. We shall also need the definition of the MGF $A_1^k(\mathbf{y})$ defined by

$$A_1^k(\mathbf{y}) := \mathrm{E}\left[\prod_{s=1}^m y_s^{R_1^{k,s}}\right], \qquad (7)$$

where

$$R_1^{k,1} = \lim_{t\to+\infty} R_1^{k,1}(t),$$
$$R_1^{k,2} = \lim_{t\to+\infty} R_1^{k,2}(t), \ldots, R_1^{k,m} = \lim_{t\to+\infty} R_1^{k,m}(t)$$

are the cumulative number of "daughter" events of first generation of type $1, 2, \ldots, m$ generated by the "mother" event of type k over all times. One may rewrite this function in probabilistic form

$$A_1^k(\mathbf{y}) := \sum_{r_1=0}^\infty \cdots \sum_{r_m=0}^\infty P_k(r_1, \ldots, r_m) \prod_{s=1}^m y_s^{r_s}, \qquad (8)$$

where $P_k(r_1, \ldots, r_m)$ is the probability that the mother event of type k generates $R^{k,1} = r_1$ first-generation events of type 1, $R^{k,2} = r_2$ first-generation events of type 2, and so on.

The events are assumed to occur after waiting times between the mother event and their occurrences that are mutually statistically independent and characterized by the probability density functions (pdf) $\{f_{k,s}(t)\}$, where all $f_{k,s}(t) \equiv 0$ for $t < 0$. Let us denote $\mathcal{P}_k(d_1, d_2, \ldots, d_m; t)$ the probability that the cumulative numbers $\{R_1^{k,s}(t)\}$ up to time t of first-generation events that have been triggered by the mother event of type k are equal to

$$R_1^{k,1}(t) = d_1, \quad R_1^{k,2}(t) = d_2, \ldots, \quad R_1^{k,m}(t) = d_m. \qquad (9)$$

Let us relate this probability $\mathcal{P}_k(d_1, d_2, \ldots, d_m; t)$ to that, denoted $\mathcal{P}_k(d_1, d_2, \ldots, d_m; t|r_1, r_2, \ldots, r_m)$, obtained under the additional condition that the total numbers of first-generation events that have been triggered by the mother event of type k over the whole time interval $t \to \infty$ are fixed at the values $\{r_1, r_2, \ldots, r_m\}$. Obviously, $\mathcal{P}_k(d_1, d_2, \ldots, d_m; t|r_1, r_2, \ldots, r_m)$ is given by a product of binomial distributions

$$\mathcal{P}_k(d_1, d_2, \ldots, d_m; t|r_1, r_2, \ldots, r_m) =$$
$$\prod_{s=1}^m \binom{r_s}{d_s} \mu_{k,s}^{d_s}(t)[1 - \mu_{k,s}(t)]^{r_s-d_s},$$
$$0 \leqslant d_1 \leqslant r_1, \ldots, \quad 0 \leqslant d_m \leqslant r_m, \qquad (10)$$

where

$$\mu_{k,s}(t) = \int_0^t f_{k,s}(t')dt'. \qquad (11)$$

Knowing the conditional probabilities $\mathcal{P}_k(d_1, d_2, \ldots, d_m; t|r_1, r_2, \ldots, r_m)$ (10), one can calculate their unconditional counterparts using the following relation:

$$\mathcal{P}_k(d_1, d_2, \ldots, d_m; t) =$$
$$\sum_{r_1=d_1}^\infty \cdots \sum_{r_m=d_m}^\infty \mathcal{P}_k(d_1, d_2, \ldots, d_m; t|r_1, r_2, \ldots, r_m)$$
$$\times P_k(r_1, r_2, \ldots, r_m), \qquad (12)$$

where $P_k(r_1, r_2, \ldots, r_m)$ is the probability that the total numbers of first-generation events that have been triggered by the mother event of type k over the whole time interval $t \to \infty$ take the values $\{r_1, r_2, \ldots, r_m\}$.

Substituting the relations (10) in (12) yields

$$\mathcal{P}_k(d_1, d_2, \ldots, d_m; t) = \sum_{r_1=d_1}^\infty \cdots \sum_{r_m=d_m}^\infty P_k(r_1, \ldots, r_m)$$
$$\times \prod_{s=1}^m \binom{r_s}{d_s} \mu_{k,s}^{d_s}(t)[1 - \mu_{k,s}(t)]^{r_s-d_s}. \qquad (13)$$

The interest in this expression (13) is that the MGF $A_1^k(y_1, y_2, \ldots, y_m; t)$ defined by expression (6) can be rewritten in probabilistic form as

$$A_1^k(\mathbf{y}; t) = \sum_{d_1=0}^\infty \cdots \sum_{d_m=0}^\infty \mathcal{P}_k(d_1, \ldots, d_m; t) \prod_{s=1}^m y_s^{d_s}. \qquad (14)$$

We are now prepared to state the following theorem, which is essential for our subsequent derivations.

Theorem 3.1. *The MGF $A_1^k(y_1, y_2, \ldots, y_m; t)$ defined by expression (6) can be represented in the form*

$$A_1^k(y_1, y_2, \ldots, y_m; t) =$$
$$Q_k[\mu_{k,1}(t)(y_1 - 1), \ldots, \mu_{k,m}(t)(y_m - 1)], \qquad (15)$$

where

$$Q_k(z_1, \ldots, z_m) := A_1^k(1 + z_1, \ldots, 1 + z_m). \qquad (16)$$

The main result can thus be rephrased explicitly as

$$A_1^k(y_1, y_2, \ldots, y_m; t) =$$
$$A_1^k(1 + \mu_{k,1}(t)(y_1 - 1), \ldots, 1 + \mu_{k,m}(t)(y_m - 1)). \qquad (17)$$

The proof is given in Appendix A.

3.2 MGF for the cumulative numbers of events over all generation triggered until time t

Let us define the MGF

$$A^k(\mathbf{y}; t) := \mathrm{E}\left[\prod_{s=1}^m y_s^{R^{k,s}(t)}\right], \qquad (18)$$

where $\{R^{k,s}(t)\}$ is the total number of events summed over all generations of events of type s triggered by a mother event of type k starting at time 0 up to time t.

Due to the branching nature of the process, the equation determining the MGF $\{A^k(\mathbf{y};t)\}$ is obtained by:

1. replacing in the left-hand-side of expression (15) $A_1^k(\mathbf{y};t)$ by $A^k(\mathbf{y};t)$; this means that we deal with events of all generations occurring till current time t;

2. replacing in the right-hand-side of expression (15) the arguments y_s by

$$y_s \Rightarrow y_s \int_0^t f_{k,s}(t'|t) A^s(\mathbf{y};t-t')dt', \qquad (19)$$

where $\{f_{k,s}(t'|t)\}$ is the conditional pdf of the random times $\{t'_{k,s}\}$ of occurrence of some first-generation event of type s triggered by the mother event of type k, under the condition that it occurred within the time interval $t' \in (0,t)$. The conditional pdf $\{f_{k,s}(t'|t)\}$ is given by

$$f_{k,s}(t'|t) = \frac{f_{k,s}(t')}{\mu_{k,s}(t)}, \qquad (20)$$

where $\mu_{k,s}(t)$ is defined by (11). The pdf $f_{k,s}(t'|t)$ inside the integral (19) takes into account that first-generation events are occurring at random times $t' < t$. The other factor $A^s(y_1,\ldots,y_m;t-t')$ takes into account all-generation events that are triggered by some first-generation event from its appearance time t' till the current time t.

The equation for the MGF $\{A^k(\mathbf{y};t)\}$ is thus

$$A^k(\mathbf{y};t) = Q_k\left[B^{k,1}(\mathbf{y};t),\ldots,B^{k,m}(\mathbf{y};t)\right], \qquad (21)$$

where

$$B^{k,s}(\mathbf{y};t) = \int_0^t f_{k,s}(t-t')[y_s A^s(\mathbf{y};t') - 1]dt'. \qquad (22)$$

3.3 General relations for the mean numbers of events over all generation triggered up to time t

The set of equations (21) together with the relations (22) provides the basis for a full description of the statistical and temporal properties of multivariate branching (Hawkes) processes. Here, we restrict our attention to the average activities, by studying the temporal dependence of the average number of events following the occurrence of a mother event of a given type.

The mean number of events of type s over all generations counted until time t that are triggered by a mother event of type k that occurred at time $t=0$, defined by

$$\bar{R}^{k,s}(t) := \mathrm{E}[R^{k,s}(t)], \qquad (23)$$

is given by the relation

$$\bar{R}^{k,s}(t) = \frac{\partial}{\partial y_s} A^k(\mathbf{y};t)\Big|_{y_1=\ldots=y_m=1}, \qquad (24)$$

where $A^k(\mathbf{y};t)$ satisfies to the set of equations (21).

In order to derive the equations determining the set $\{\bar{R}^{k,s}(t)\}$, we need to state the following properties exhibited by the functions $Q_k(\mathbf{y})$ given by (16) and $B^{k,s}(\mathbf{y};t)$

given by (22), which are contributing to equation (21). From the definition of the functions $B^{k,s}(\mathbf{y};t)$ and of the MGF $A^k(\mathbf{y};t)$, we have

$$B^{k,s}(\mathbf{y};t)\big|_{y_1=\ldots=y_m=1} \equiv 0 \qquad (25)$$

and

$$\frac{\partial}{\partial y_s} B^{k,\ell}(\mathbf{y};t,\tau)\big|_{y_1=\ldots=y_m=1} = \mu_{k,s}(t)\delta_{\ell,s} + f_{k,\ell}(t) \otimes \bar{R}^{\ell,s}(t). \qquad (26)$$

The convolution operation is defined as usual by $f(t) \otimes g(t) := \int_0^t f(t-t')g(t')dt'$. Moreover, the following equality holds

$$\frac{\partial}{\partial y_s} Q_k(\mathbf{y};t,\tau)\big|_{y_1=\ldots=y_m=0} = n_{k,s}, \qquad (27)$$

where $n_{k,s}$ is the mean value of the total number of first-generation events of type s triggered by a mother of type k. The set $\{n_{k,s}\}$ for all k's and s's generalize the average branching ratio n defined by expression (4) above for monovariate branching processes and are given by

$$n_{k,s} = \frac{\partial}{\partial y_s} A_1^k(\mathbf{y})\big|_{y_1=\ldots=y_m=1} \qquad (28)$$

where $A_1^k(\mathbf{y})$ is defined in (7).

Using the above relations (25)–(27), the equality (24) together with equation (21) yields

$$\bar{R}^{k,s}(t) = n_{k,s} \cdot \mu_{k,s}(t) + \sum_{\ell=1}^m n_{k,\ell} f_{k,\ell}(t) \otimes \bar{R}^{\ell,s}(t). \qquad (29)$$

Introducing the event rates, i.e., the number of events per unit time,

$$\rho^{k,s}(t) = \frac{d\bar{R}^{k,s}(t)}{dt}, \qquad (30)$$

expression (29) transforms into

$$\rho^{k,s}(t) = n_{k,s}(t) + \sum_{\ell=1}^m n_{k,\ell}(t) \otimes \rho^{\ell,s}(t), \qquad (31)$$

where we have used the following notation

$$n_{k,s}(t) := n_{k,s} \cdot f_{k,s}(t). \qquad (32)$$

The set of equations (31) for all k's and s's constitute the fundamental starting point of our analysis.

In order to make further progress, in view of the convolution operator, it is convenient to work with the Laplace transform of the event rates:

$$\tilde{\rho}^{k,s}(u) = \int_0^\infty \rho^{k,s}(t)e^{-ut}dt. \qquad (33)$$

Introducing the matrices

$$\tilde{\Phi}(u) = [\tilde{\rho}^{k,s}(u)] \quad \text{and} \quad \tilde{N}(u) = [\tilde{n}_{k,s}(u)], \qquad (34)$$

we obtain the following equation for the matrix $\tilde{\Phi}(u)$:

$$\hat{I}\,\tilde{\Phi}(u) = \tilde{N}(u) + \tilde{N}(u)\,\tilde{\Phi}(u), \qquad (35)$$

whose solution is

$$\tilde{\Phi}(u) = \frac{\tilde{N}(u)}{\hat{I} - \tilde{N}(u)}. \tag{36}$$

The rest of the paper is devoted to the study of this solution (36) for various system structures and memory kernels.

4 Symmetric mutual excitations

4.1 Definitions

Let us consider the case where the set $\{n_{k,s}\}$ defined by expression (28) reduces to

$$n_{k,k} = a; \qquad n_{k,s} = b, \quad k \neq s. \tag{37}$$

This form (37) means that the efficiency to generate an event of a specific type depends only on whether the 'mother' and 'daughter' have the same type. As a consequence, the mean number of first-generation events of all kinds that are triggered by a mother event of some type k,

$$n_k = \sum_{s=1}^{m} n_{k,s}, \tag{38}$$

is independent of k and given by

$$n_k = n = a + (m-1)b, \quad \text{for all } k. \tag{39}$$

Introducing the factor

$$q = \frac{b}{a} \tag{40}$$

comparing the inter-types with the intra-type triggering efficiencies, we obtain

$$a = \frac{n}{1 + (m-1)q}, \qquad b = \frac{nq}{1 + (m-1)q}. \tag{41}$$

In the time domain, we consider the case of symmetric mutual excitations such that all pdf's $f_{k,s}(t) \equiv f(t)$ are independent of the indexes k and s and are all equal to each other. There is no contradiction with this assumption and the fact that there are two coefficients a and b characterizing event triggering, since the time-dependent functions $f_{k,s}(t)$ are normalized to 1.

4.2 General solution in terms of Laplace transforms

With the definitions of Section 4.1, it follows that $\tilde{n}_{k,s}(u) = n\tilde{f}(u)$ and one can show that all diagonal and non-diagonal entries of the matrix $\tilde{\Phi}(u)$ given by (36) are

given, respectively by

$$
\begin{aligned}
\tilde{\rho}(u) &:= \tilde{\rho}^{k,k}(u) \\
&= \frac{n\tilde{f}(u)}{1 - n\tilde{f}(u)} \frac{1 + n\tilde{f}(u)(q-1)}{1 + n\tilde{f}(u)(q-1) + q(m-1)}, \\
\tilde{g}(u) &:= \tilde{\rho}^{k,s}(u) \\
&= \frac{n\tilde{f}(u)}{1 - n\tilde{f}(u)} \frac{q}{1 + n\tilde{f}(u)(q-1) + q(m-1)}, \\
&\hspace{6cm} k \neq s. \quad (42)
\end{aligned}
$$

Moreover, the Laplace transform

$$\tilde{\rho}^k(u) = \sum_{s=1}^{m} \tilde{\rho}^{k,s}(u) = \tilde{\rho}(u) + (m-1)\tilde{g}(u) \tag{43}$$

of the total rate of events of all types triggered by a mother jump of type k defined by $\rho^k(t) = \sum_{s=1}^{m} \rho^{k,s}(t)$ satisfies the relation

$$\tilde{\rho}^k(u) = \frac{n\tilde{f}(u)}{1 - n\tilde{f}(u)}. \tag{44}$$

4.3 Exponential pdf of triggering times of first-generation events

Let us first study the case where the pdf $f(t)$ of the waiting times to generate first-generation events is exponential:

$$f(t) = \alpha e^{-\alpha t} \iff \tilde{f}(u) = \frac{\alpha}{\alpha + u}. \tag{45}$$

The inverse Laplace transforms of the solutions (42) are then

$$
\begin{aligned}
\rho(t) &:= \rho^{k,k}(t) = \alpha \frac{n}{m} e^{-(1-n)\alpha t} \left(1 + (m-1)\gamma e^{-n(1-\gamma)\alpha t} \right), \\
g(t) &:= \rho^{k,s}(t) = \alpha \frac{n}{m} e^{-(1-n)\alpha t} \left(1 - \gamma e^{-n(1-\gamma)\alpha t} \right), \quad k \neq s,
\end{aligned}
\tag{46}
$$

where

$$\gamma := \gamma(q,m) = \frac{1-q}{1 + (m-1)q}. \tag{47}$$

Expressions (46) give the explicit time-dependence of two functions $\rho(t)$ and $g(t)$:

- $\rho(t) := \rho^{k,k}(t)$ is the rate of events over all generations of some type k resulting from a given mother event of the same type k. Notice that the term "over all generations" means that an event of type k occurring at some time $t > 0$, and belonging to the descent of some previous mother of the same type k that occurred at time 0, may have been generated through a long cascade of intermediate events of possibly different kinds, via random changes of types along the genealogy.
- $g(t) := \rho^{k,s}(t)$ is the rate of events over all generations of some type s resulting from a given mother event of a different type k. As for $\rho(t)$, an event of type s occurring at some time $t > 0$, and belonging to the descent
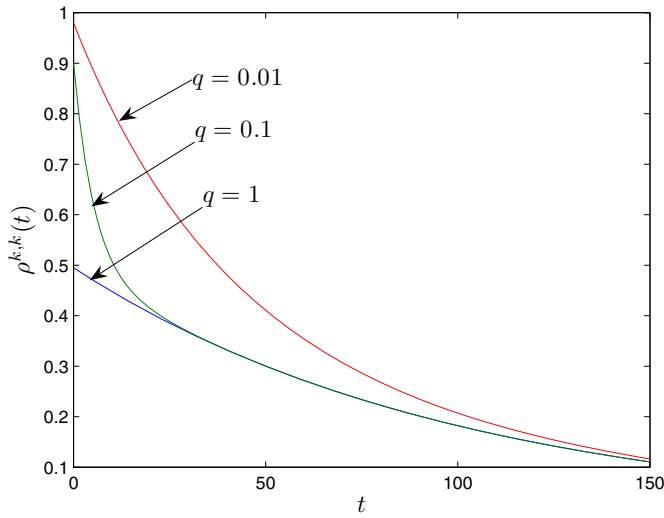
Fig. 5. Time-dependence of $\rho(t) := \rho^{k,k}(t)$, for $m = 3$, $n = 0.99$, and $q = 1; 0.1; 0.01$. Time is in unit of $1/\alpha$.
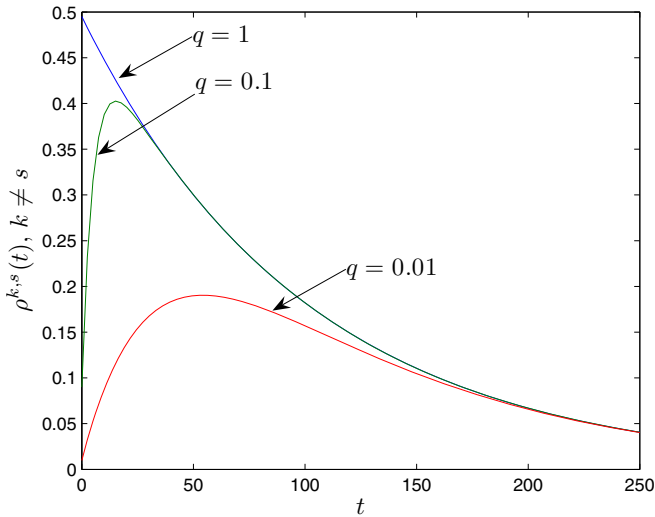

Fig. 6. Time-dependence of $g(t) := \rho^{k,s}(t)$, for $m = 3$, $n = 0.99$ and $q = 1; 0.1; 0.01$. Time is in unit of $1/\alpha$.

of some previous mother of a different type k that occurred at time 0, may have been generated through a long cascade of intermediate events of possibly different kinds, via random changes of types along the genealogy.

Figures 5 and 6 show, respectively, $\rho(t) := \rho^{k,k}(t)$ and $g(t) := \rho^{k,s}(t)$ for the case of $m = 3$ types of events and rather close to criticality ($n = 0.99$), for different coupling amplitudes $q = 1; 0.1; 0.01$.

The case $q = 1$ of complete coupling is special in two ways: (i) $\rho(t)$ and $g(t)$ are proportional to each other; (ii) there is only one time scale $\tau_1 = \frac{1}{\alpha}\frac{1}{1-n}$. In contrast, as soon as $q < 1$, i.e., events of a given type tends to trigger more events of the same type than events of different types, one can observe that the time-dependence of $\rho(t)$ and $g(t)$ become qualitatively different. The monotonous decay of $\rho(t)$ can be contrasted with the non-monotonous bell-shape dependence of $g(t)$. This non-monotonous be-

havior of $g(t)$ results from the progressive seeding of events of a different type than the initial mother type by the less efficient mutual excitation process. This is associated with the introduction of a second time scale $\tau_2 = \frac{1}{\alpha}\frac{1}{1-n\gamma} \leq \tau_1$ controlling the dynamics of both $\rho(t)$ and $g(t)$ at short times. This effect is all the stronger, the smaller q is, i.e., the larger γ is.

This phenomenon of the occurrence of a second time scale τ_2 and of the distinct behavior of $g(t) := \rho^{k,s}(t)$ for $q < 1$ constitutes a characteristic signature of the mutual excitation mechanism, in a system in which the time response function is exponential (Poisson, i.e., without memory). The cascade of triggering together with the random changes of types of events break the Poissonian nature of the relaxation of the event activity triggered by a given mother ancestor. This is different from what occurs for a single event type $m = 1$ for which the existence of multiple generations do not change the Poissonian nature of the relaxation process. It only extends the time scale according to $\tau_1 = \frac{1}{\alpha}\frac{1}{1-n}$ as the average branching ratio n increases to the critical value 1.

4.4 Power law pdf of triggering times of first-generation events

The same qualitative picture emerges for other pdf's such as power laws, with a monotonous decay of $\rho(t)$ coexisting with a growth from zero up to a maximum followed by a decay for $g(t)$. But more interesting features appear, such as the renormalization of the exponents in two distinct families, as we now show. As many systems exhibit power law pdf's of waiting times with rather small exponents $\theta \leq 0.5$ (see definition in expression (48)), we shall consider this regime in the following.

For definiteness, we consider the pdf with power law tail given by

$$f(t) = \frac{\alpha\theta}{(1+\alpha t)^{1+\theta}}, \qquad t > 0, \qquad \theta > 0. \qquad (48)$$

The constant 1 in the denominator regularizes the pdf at times $t < 1/\alpha$. The corresponding Laplace transform of (48) is

$$\tilde{f}(u) = \theta\, e^{u/\alpha}\left(\frac{u}{\alpha}\right)^\theta \Gamma\left(-\theta, \frac{u}{\alpha}\right), \qquad (49)$$

which can then be used in (42) to get the general solutions.

4.4.1 Non-critical same-type activity rate $\rho(t) =: \rho^{k,k}(t)$

We first rewrite the Laplace transform $\tilde{\rho}(u)$ given by (42) in a form more convenient for its analysis:

$$\tilde{\rho}(u) = \bar{R}^{k,k}\frac{1-\varphi(u)}{1+\gamma_0\varphi(u)}\frac{1+\gamma_1\varphi(u)}{1+\gamma_2\varphi(u)}, \qquad (50)$$

where

$$\varphi(u) := 1 - \tilde{f}(u) \qquad (51)$$

and

$$\gamma_0 = \frac{n}{1-n}, \quad \gamma_1 = \frac{n(1-q)}{1-n+qn},$$

$$\gamma_2 = \frac{n(1-q)}{1-n+q(n+m-1)}, \gamma_0 > \gamma_1 > \gamma_2 \quad (q > 0, m > 1). \tag{52}$$

The long time behavior of $\rho(t)$ is controlled by the small u properties of $\tilde{\rho}(u)$, itself dependent on the behavior of $\varphi(u)$ for small u. From its definition (51), we have that $\varphi(u) \to 0$ for $u \to 0$. Then, the asymptotic behavior of the Laplace transform (50) of $\rho(t)$ is

$$\tilde{\rho}(u) \sim \bar{R}^{k,k} - \bar{R}^{k,k}\gamma_\rho\varphi(u),$$

$$\gamma_\rho = 1 + \gamma_0 + \gamma_2 - \gamma_1. \tag{53}$$

For $\theta \in (0,1)$, the auxiliary function $\varphi(u)$ (51) has the following asymptotic behavior

$$\varphi(u) \sim \beta v^\theta \ll 1, \quad v \ll 1,$$

$$\beta = \Gamma(1-\theta), \quad v = \frac{u}{\alpha}. \tag{54}$$

Accordingly, relation (53) transforms into

$$\tilde{\rho}(u) \sim \bar{R}^{k,k}\left[1 - \gamma_\rho\beta v^\theta\right]. \tag{55}$$

The corresponding asymptotic of the rate $\rho(t)$ is thus

$$\rho(t) \sim \bar{R}^{k,k}\frac{\gamma_\rho\beta\theta}{\Gamma(1-\theta)}\frac{1}{(\alpha t)^{1+\theta}}$$

$$= \bar{R}^{k,k}\gamma_\rho\theta\frac{1}{(\alpha t)^{1+\theta}}, \quad t \to \infty. \tag{56}$$

The rate $\rho(t) := \rho^{k,k}(t)$ of events over all generations of some type k resulting from a given mother event of the same type k that occurred at time 0 decays with the same power law behavior as the bare memory function or pdf of waiting times for first-generation events. The only significant difference is the renormalization of the amplitude by the factor $\bar{R}^{k,k}\gamma_\rho$ resulting from the cascades of generations and random changes between the different event types.

4.4.2 Intermediate critical asymptotic same-type activity rate $\rho(t) := \rho^{k,k}(t)$

For n close to 1, γ_0 becomes large and there is an interesting intermediate asymptotic regime describing the intermediate time decay of $\rho(t)$. To describe it, we need to distinguish the following three parameter regimes.

1. $\gamma_0 \gg 1$ and $\gamma_2 < \gamma_1 \simeq 1$. This occurs for $n \to 1$ while q is not too close to 0. For instance, $n = 0.95$, $q = 0.5$, $m = 5$ yield $\gamma_2 = 0.613, \gamma_1 = 0.905, \gamma_0 = 19$. In this case, there is an intermediate range on the u-axis defined by

$$(\gamma_0\beta)^{-\theta} \ll v \ll (\gamma_1\beta)^{-\theta}$$

$$\iff \alpha(\gamma_0\beta)^{-\theta} \ll u \ll \alpha(\gamma_1\beta)^{-\theta} \approx \alpha, \tag{57}$$

such that $\gamma_0\varphi(u) \gg 1$ while $\gamma_1\varphi(u) \ll 1$ and $\gamma_2\varphi(u) \ll 1$. In this range, the leading terms controlling the value of expression (50) is

$$\tilde{\rho}(u) \approx \bar{R}^{k,k}\frac{1}{\gamma_0\varphi(u)}. \tag{58}$$

Substituting the asymptotic relation (54), we obtain

$$\tilde{\rho}(u) \approx \bar{R}^{k,k}\frac{1}{\gamma_0\Gamma(1-\theta)} v^{-\theta}. \tag{59}$$

The corresponding intermediate asymptotic of the rate is

$$\rho(t) \approx \bar{R}^{k,k}\frac{1}{\gamma_0}\frac{\sin(\pi\theta)}{\pi}\frac{1}{(\alpha t)^{1-\theta}},$$

$$1 \ll \alpha t \ll [\gamma_0\Gamma(1-\theta)]^{1/\theta}. \tag{60}$$

2. $\gamma_0 > \gamma_1 \gg 1$ and $\gamma_2 \simeq 1$: this occurs for $n \to 1$ with q close to 0 and m large. For instance, $n = 0.95$, $q = 0.01$, $m = 50$ yield $\gamma_2 = 1.68, \gamma_1 = 15.8, \gamma_0 = 19$. In this case, there is an intermediate range on the u-axis defined by

$$(\gamma_1\beta)^{-\theta} \ll v \ll (\gamma_2\beta)^{-\theta}$$

$$\iff \alpha(\gamma_1\beta)^{-\theta} \ll u \ll \alpha(\gamma_2\beta)^{-\theta} \approx \alpha, \tag{61}$$

such that $\gamma_0\varphi(u) > \gamma_1\varphi(u) \gg 1$ while $\gamma_2\varphi(u) \ll 1$. In this range, the leading terms controlling the value of expression (50) is

$$\tilde{\rho}(u) = \bar{R}^{k,k}\frac{\gamma_1}{\gamma_0}(1 - \varphi(u)), \tag{62}$$

whose inverse Laplace transform has the same form as (56) with γ_ρ replaced by 1.

3. $\gamma_0 > \gamma_1 > \gamma_2 \gg 1$: this occurs for $n \to 1$ with q close to 0 and m not too large. For instance, $n = 0.95$, $q = 0.01$, $m = 2$ yield $\gamma_2 = 13.5, \gamma_1 = 15.8, \gamma_0 = 19$. Then, in the intermediate interval on the u-axis

$$(\gamma_2\beta)^{-\theta} \ll v \ll 1 \iff \alpha(\gamma_2\beta)^{-\theta} \ll u \ll \alpha, \tag{63}$$

the asymptotic relation (54) holds, while at the same time $\gamma_2\varphi(u) \gg 1$. In this range (63), the leading terms controlling the value of expression (50) is

$$\tilde{\rho}(u) \approx \bar{R}^{k,k}\frac{1}{\gamma_0\varphi(u)}\frac{\gamma_1\varphi(u)}{\gamma_2\varphi(u)} = \bar{R}^{k,k}\frac{\gamma_1}{\gamma_0\gamma_2}\frac{1}{\varphi(u)}. \tag{64}$$

Using the asymptotic relation (54), we obtain

$$\tilde{\rho}(u) \approx \mathcal{G} v^{-\theta},$$

$$\mathcal{G} = \bar{R}^{k,k}\frac{\gamma_1}{\gamma_0\gamma_2\beta} = \bar{R}^{k,k}\frac{\gamma_1}{\gamma_0\gamma_2\Gamma(1-\theta)}. \tag{65}$$

The corresponding intermediate power asymptotic of the rate $\rho(t)$ is

$$\rho(t) \approx \bar{R}^{k,k}\frac{\gamma_1}{\gamma_0\gamma_2}\frac{\sin(\pi\theta)}{\pi}\frac{1}{(\alpha t)^{1-\theta}},$$

$$1 \ll \alpha t \ll [\gamma_2\Gamma(1-\theta)]^{1/\theta}. \tag{66}$$

As an illustration, for the above values $n = 0.95$, $q = 0.01$, $m = 2$, the power law (66) with exponent $1 - \theta$ holds up to a maximum time $[\gamma_2 \Gamma(1-\theta)]^{1/\theta} \alpha^{-1} \approx 75\,000\,\alpha^{-1}$ for $\theta = 0.25$.

Let us summarize and interpret the above results:

1. $n \to 1$ and q not small ($\gamma_0 \gg 1$ and $\gamma_2 < \gamma_1 \simeq 1$). The intermediate power law asymptotic (60) with exponent $1 - \theta$ is similar to the renormalized response function due to the cascade of generations found for the self-excited Hawkes process with just one type of events [11,65,71]. The mechanism is the same, since a coupling coefficient q not too small ensures a good mixing among all generations.

2. $n \to 1$ with $q \to 0$ and $m \to \infty$ with $qm \simeq 1$ ($\gamma_0 > \gamma_1 \gg 1$ and $\gamma_2 \simeq 1$). In contrast with the previous case, the activity rate $\rho(t)$ exhibits the same decay (56) with exponent $1 + \theta$ as if the system was far from criticality. In a sense, due to the weak mutual triggering efficiency and the many event types, the system is never critical.

3. $n \to 1$ with $q \to 0$ and m not too large such that $qm \ll 1$ ($\gamma_0 > \gamma_1 > \gamma_2 \gg 1$). The intermediate power law asymptotic (66) with exponent $1-\theta$ is again similar to the renormalized response function due to the cascade of generations found for the self-excited Hawkes process with just one type of events [11,65,71].

4.4.3 Asymptotic and intermediate critical asymptotic inter-type activity rate g(t) := $\rho^{k,s}$(t)

In order to obtain the time-dependence of $g(t)$, we express its Laplace transform $\tilde{g}(u)$ given in expression (42) in a form analogous to (50):

$$\tilde{g}(u) = \bar{R}^{k,s} \frac{1 - \varphi(u)}{1 + \gamma_0 \varphi(u)} \frac{1}{1 + \gamma_2 \varphi(u)}, \quad k \neq s. \qquad (67)$$

Three regimes can be distinguished:

1. **Asymptotic regime of long times for $n < 1$.** At long times, the asymptotic relation (54) holds true. Then, analogous to (55) and (56), we obtain the following asymptotics

$$\tilde{g}(u) \sim \bar{R}^{k,s} \left[1 - \gamma_g \beta v^\theta\right], \gamma_g = 1 + \gamma_0 + \gamma_2 \quad u \to 0$$
$$\Rightarrow g(t) \sim \bar{R}^{k,s} \frac{\gamma_g \beta \theta}{\Gamma(1-\theta)} \frac{1}{(\alpha t)^{1+\theta}}, \quad t \to \infty. \quad (68)$$

This power law decay with exponent $1 + \theta$, equal to the exponent of the memory kernel (48), is characteristic of the non-critical regime in which only a few generations of events are triggered in significant numbers.

2. **Intermediate asymptotic regime ($n \to 1$ with $q \to 0$ and m large).** Then, $\gamma_0 \gg 1$ and $\gamma_2 \simeq 1$. This is the same second regime analyzed in Section 4.4.2. In this case, there is an intermediate asymptotic in the range defined by (61) such that the following approximate relation holds

$$\tilde{g}(u) \approx \frac{\bar{R}^{k,s}}{\gamma_0} \frac{1}{\varphi(u)}, \qquad k \neq s. \qquad (69)$$

The corresponding intermediate power asymptotic of $g(t) := \rho^{k,s}(t)$ is

$$\rho(t) \approx \frac{\bar{R}^{k,k}}{\gamma_0} \frac{\sin(\pi\theta)}{\pi} \frac{1}{(\alpha t)^{1-\theta}}, \quad 1 \ll \alpha t \ll [\gamma_0 \Gamma(1-\theta)]^{1/\theta}. \tag{70}$$

This power law decay with exponent $1 - \theta$ is significantly slower than the previous one with exponent $1 + \theta$ and results from the proximity to the critical point $n = 1$.

3. **Intermediate asymptotic regime ($n \to 1$ with $q \to 0$ and m small).** Then, γ_2 given in (52) is large and there is an intermediate interval (63) for u such that, analogous to (64), the following approximate relation holds

$$\tilde{g}(u) \approx \bar{R}^{k,s} \frac{1}{\gamma_0 \varphi(u)} \frac{1}{\gamma_2 \varphi(u)} = \bar{R}^{k,s} \frac{1}{\gamma_0 \gamma_2} \frac{1}{\varphi^2(u)}. \quad (71)$$

Using the power asymptotic (54), we obtain the intermediate power law

$$\tilde{g}(u) \approx \mathcal{G} \; v^{-2\theta}, \; \mathcal{G} = \bar{R}^{k,s} \frac{1}{\gamma_0 \gamma_2 \beta} = \bar{R}^{k,s} \frac{1}{\gamma_0 \gamma_2 \Gamma(1-\theta)}. \tag{72}$$

Accordingly, the intermediate power law asymptotic of the inter-type activity rate $g(t) := \rho^{k,s}(t)$ reads

$$g(t) \approx \frac{\bar{R}^{k,s}}{\gamma_0 \gamma_2 \Gamma(2\theta) \Gamma(1-\theta)} \frac{1}{(\alpha t)^{1-2\theta}},$$
$$1 \ll \alpha t \ll [\gamma_2 \Gamma(1-\theta)]^{1/\theta}. \quad (73)$$

This power law decay with exponent $1 - 2\theta$ is similar to the decay after an "endogenous" peak, as classified in previous analyses of the monovariate self-excited Hawkes process [56,67–70,72]. Indeed, the exponent $1 - 2\theta$, corresponding to a very slow power law decay, has been until now seen as the characteristic signature of self-organized bursts of activities that are generated endogenously without the need for a major exogenous shock. Here, we see this exponent describing the decay of the activity of events triggered by an "exogenous" mother shock of a different type, in the critical regime $n \to 1$ and for weak mutual coupling $q \to 0$. It is clear that the mechanism is different from the previously classified "endogenous" channel [56,67–70,72], involving here an interplay between the cascade over generations and the weak mutual excitations.

Figure 7 shows the time-dependence of both $\rho(t) := \rho^{k,k}(t)$ and $g(t) := \rho^{k,s}(t)$ in the regime $n \to 1$ with $q \to 0$ and m small, for which there exists an intermediate asymptotic of the third type both for $\rho(t)$ and $g(t)$. For $\rho(t)$, one can clearly observe the intermediate power law asymptotic with exponent $1-\theta = 0.75$ followed by the final asymptotic power law with exponent $1 + \theta = 1.25$. For $g(t)$, the intermediate power law asymptotic with exponent $1 - 2\theta = 0.5$ is clearly observed, followed by the same final asymptotic power law with exponent $1 + \theta = 1.25$.
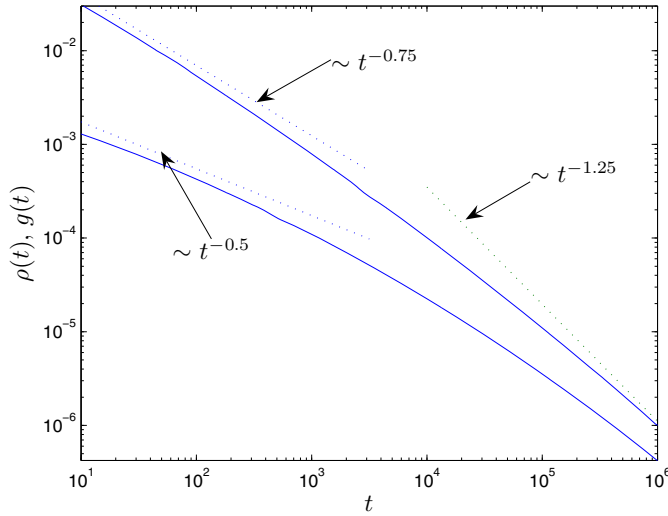
Fig. 7. Time-dependence of $\rho(t) := \rho^{k,k}(t)$ and $g(t) := \rho^{k,s}(t)$ in double logarithmic representation for $n = 0.95$, $q = 0.01$, $m = 2$, and $\theta = 0.25$ ($1 + \theta = 1.25$, $1 - \theta = 0.75$, $1 - 2\theta = 0.5$). Time is in unit of α^{-1}. The functions $\rho(t)$ and $g(t)$ have been calculated numerically from their complete Laplace transforms (42).

5 One-dimensional chain of directed triggering

5.1 Definitions

We consider a chain of directed influences $k \to k+1$ where the events of type k trigger events of both types k and $k+1$ only (and not events of type $k-1$ or any other types), and this for $k = 1, 2, \ldots, m$. This is captured by a form of the matrix $\hat{N}$ which has only the diagonal and the line above the diagonal with non-zero elements.

As the simplest example, we shall study networks of mutual excitations corresponding to the following matrix $\hat{N}$ of the mean numbers of first-generation events

$$\hat{N} = \begin{bmatrix} \chi & \xi & 0 & 0 & 0 & \ldots & 0 & \ldots \\ 0 & \chi & \xi & 0 & 0 & \ldots & 0 & \ldots \\ 0 & 0 & \chi & \xi & 0 & \ldots & 0 & \ldots \\ 0 & 0 & 0 & \chi & \xi & \ldots & 0 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \ldots \end{bmatrix} \qquad (74)$$

where

$$\chi = \frac{n}{1+q}, \quad \xi = \frac{nq}{1+q}. \qquad (75)$$

5.2 Laplace transform of the event activities $\rho^{k,s}(t)$ defined in (30)

In order to derive the equations governing the rates $\rho^{k,s}(t)$, we need to recall a result concerning the total numbers of events $\bar{R}^{k,k}$ and $\bar{R}^{k,s}$ generated by a given mother of type k.

The statistical average of the total numbers of events of type s over all generations that are triggered by a mother

of type k is given by

$$\bar{R}^{k,s} = \frac{\partial}{\partial y_s} A^k(\mathbf{y}) \Big|_{y_1 = y_2 = \ldots = y_m = 1}. \qquad (76)$$

It is standard in branching processes to find that the generating functions $\{A^k(\mathbf{y}); k = 1, \ldots, m\}$ are solutions of the transcendent equations (see Ref. [73] for details of the derivation):

$$A^k(\mathbf{y}) = A_1^k(y_1 \cdot A^1(\mathbf{y}), \ldots, y_m \cdot A^m(\mathbf{y})), \ k = 1, \ldots, m. \qquad (77)$$

Using (77), it is straightforward to show that $\bar{R}^{k,s}$ is solution of

$$\bar{R}^{k,s} = n_{k,s} + \sum_{\ell=1}^{m} n_{k,\ell} \cdot \bar{R}^{\ell,s}, \qquad (78)$$

where $n_{k,s}$ is the mean number of first-generation events of type s triggered by the mother event of type k.

Since expression (78) holds for all $k = 1, \ldots, m$ and $s = 1, \ldots, m$, it can be written in matrix form

$$\hat{R} = \hat{N} + \hat{N}\hat{R}, \qquad (79)$$

where $\hat{N} = [n_{k,s}]$ is the matrix of the mean numbers of first-generation events and $\hat{R} = [\bar{R}^{k,s}]$ is the matrix of the mean numbers of events over all generations. The sum over row indices of the elements of the matrix $\hat{N}$

$$n_k = \sum_{s=1}^{m} n_{k,s}, \qquad (80)$$

is the mean number of first-generation events of all kinds that are triggered by a mother event of type k.

The solution of the matrix equation (79) is

$$\hat{R} = \frac{\hat{N}}{\hat{I} - \hat{N}}. \qquad (81)$$

Assuming that the mother event is of type k, we thus obtain [73] $\bar{R}^{s,s} = \bar{R}^{k,s} = \bar{R}^{s,k} = 0$ for $1 \le s < k$ and

$$\begin{aligned} \bar{R}^{k,k} &= \frac{\chi}{1-\chi} = \frac{n}{1+q-n}, \\ \bar{R}^{k,s} &= \frac{\xi^{s-k}}{(1-\chi)^{s-k+1}} \\ &= (1+q)\frac{(nq)^{s-k}}{(1+q-n)^{s-k+1}}, \quad s > k. \end{aligned} \qquad (82)$$

Then, the Laplace transforms $\tilde{\rho}^{k,s}(u)$ (33) of the rates $\rho^{k,s}(t)$ are given by the right-hand sides of expressions (82), with the following substitutions

$$\chi \mapsto \chi\tilde{f}(u), \quad \xi \mapsto \xi\tilde{f}(u). \qquad (83)$$

This yields

$$\begin{aligned} \tilde{\rho}^{k,k}(u) &:= \tilde{\rho}(u) = \frac{\chi\tilde{f}(u)}{1 - \chi\tilde{f}(u)}, \\ \rho^{k,s}(u) &:= \tilde{g}_m(u) = \frac{(\xi\tilde{f}(u))^m}{(1 - \chi\tilde{f}(u))^{m+1}}, \quad m = s - k > 0. \end{aligned} \qquad (84)$$
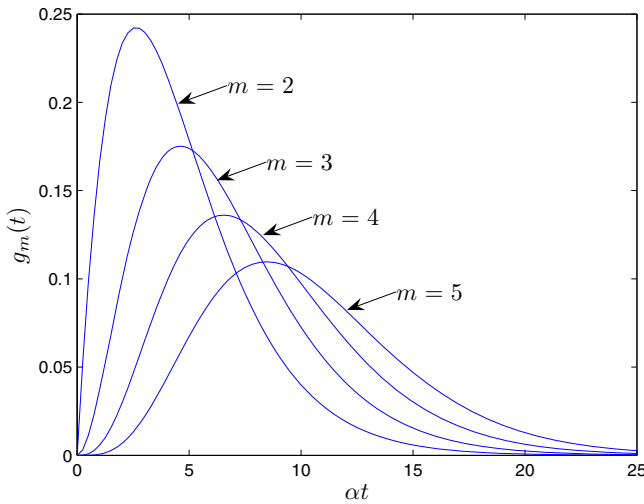
Fig. 8. Time-dependence of the cross-rates $\rho^{k,s}(t) := g_m(t)$ given by (85) for $n = 0.95$, $q = 1$ and different m values.

5.3 Exponential pdf f(t) of triggering times of first-generation events

We use the parameterization (45) for the pdf $f(t)$, which leads after calculations to

$$
\begin{aligned}
\rho^{k,k}(t) &:= \rho(t) = \alpha\chi \ e^{-(1-\chi)\alpha t}, \\
\rho^{k,s}(t) &:= g_m(t) = \frac{\alpha\xi}{m!}(\xi\alpha t)^{m-1}(m + \chi\alpha t) \ e^{-(1-\chi)\alpha t}, \\
&\qquad\qquad\qquad m = s - k > 0.
\end{aligned}
\tag{85}
$$

The rate $\rho^{k,k}(t) := \rho(t)$ of events of the same type as the mother decays simply as an exponential with a characteristic decay time $\frac{1+q}{1+q-n}\alpha^{-1}$, which exhibits the standard critical slowing down at the critical value of the mean branching ratio $n_c = 1 + q$. In contrast, the cross rates $\rho^{k,s}(t) := g_m(t)$ exhibit a non-monotonous behavior, which reflects the directed nature of the mutual triggering of events of different types. For large m values, the cross-rate becomes almost symmetrical functions of time, as shown in Figure 8.

5.4 Power law pdf of triggering times of first-generation events

The time-dependence of $\rho^{k,k}(t) := \rho(t)$ given by the first equation of (84) when $f(t)$ is given by relation (48) with Laplace transform (49) is the same as for the monovariate Hawkes process (with a single event type), with the modification that the role of the mean branching ratio n is replaced by χ. The rate $\rho(t)$ exhibits an intermediate power law asymptotic with exponent $1 - \theta$ up to a cross-over time $\simeq \alpha^{-1}/(1-\chi)^{\frac{1}{\theta}}$ followed by the asymptotic power law decay with exponent $1 + \theta$ corresponding to the memory kernel $f(t)$.

Interesting new regimes appear for the time-dependences of the cross-rates $g_m(t)$. We first express the

Laplace transforms $\tilde{g}_m(u)$ given by (84) of the cross-rates by using the auxiliary function $\varphi(u)$ defined by (51):

$$
\tilde{g}_m(u) = \bar{R}^{k,s} \frac{[1 - \varphi(u)]^m}{[1 + \gamma\varphi(u)]^{m+1}},
\tag{86}
$$

where

$$
\gamma = \frac{\chi}{1-\chi} = \frac{n}{1+q-n},
\tag{87}
$$

and the mean number $\bar{R}^{k,s}$ ($k \neq s$) is given by expression (82). Replacing $\varphi(u)$ by its asymptotic (54), we obtain the asymptotic formula

$$
\tilde{g}_m(u) = \bar{R}^{k,s} \frac{\left(1 - \beta v^\theta\right)^m}{\left(1 + \gamma\beta v^\theta\right)^{m+1}}.
\tag{88}
$$

The long time asymptotic of $g_m(t)$ is controlled by the behavior of $\tilde{g}_m(u)$ for $v \to 0$, whose leading order is given by

$$
\begin{aligned}
\tilde{g}_m(u) &\sim \bar{R}^{k,s} \left[1 - \gamma(m)v^\theta\right], \qquad v \to 0 \\
\gamma(m) &= \beta(m + (m+1)\gamma).
\end{aligned}
\tag{89}
$$

This expression holds true for $\gamma < +\infty$, i.e., $n < 1 + q$, where the upper bound $n_c = 1 + q$ define the critical point. Accordingly, the main asymptotic of the cross event rate $g_m(t)$ is

$$
g(t) \sim \bar{R}^{k,s} \frac{\gamma(m)\theta}{\Gamma(1-\theta)} \frac{1}{(\alpha t)^{1+\theta}}, \quad t \to \infty.
\tag{90}
$$

This recovers the usual long time power law dependence, which is determined by the memory kernel $f(t)$ of waiting times for first-generation triggering.

There is also an intermediate asymptotic regime present when $\gamma \gg 1$, i.e., $n \to 1 + q$ from below. Specifically, the intermediate asymptotic domain in the variable v is defined by the interval $(\beta\gamma)^{-1/\theta} \ll v \ll 1$, such that $\gamma\beta v^\theta \gg 1$ while $v \ll 1$. Then, the asymptotic relation (54) is true, and expression (88) can be simplified into the approximate relation

$$
\tilde{g}_m(u) \approx \mathcal{G}_m v^{-(m+1)\theta}, \quad \mathcal{G}_m = \frac{\bar{R}^{k,s}}{(\gamma\beta)^{m+1}},
$$
$$
(\beta\gamma)^{-1/\theta} \ll v \ll 1. \tag{91}
$$

Accordingly, analogous to (73), we obtain

$$
g_m(t) \approx \frac{\mathcal{G}_m}{\Gamma[(m+1)\theta]} \frac{1}{(\alpha t)^{1-(m+1)\theta}}, \quad 1 \ll \alpha t \ll (\gamma\beta)^{1/\theta}.
\tag{92}
$$

This expression (92) predicts a hierarchy of exponents $1 - (m + 1)\theta$ characterizing the intermediate asymptotic power law dependence of the rates $g_m(t) := \rho^{k,s}(t)$ of events of type s as a function of the distance $m = s - k$ along the space of types from the type k of the initial triggering mother. Figure 9 illustrates this prediction (92) for $m = 1; 2$ with $\theta = 0.25$, leading to the two intermediate asymptotic exponents $1 - 2\theta = 0.5$ and $1 - 3\theta = 0.25$.
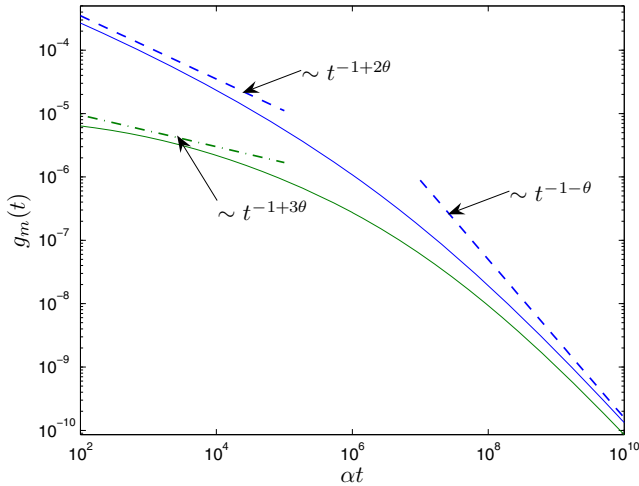
Fig. 9. Time-dependence of the rates $g_1(t) := \rho^{k,k+1}(t)$ (upper curve) and $g_2(t) := \rho^{k,k+2}(t)$ (lower curve) of events of type $s = k+1$ and $s = k+2$ triggered by a mother of type k. The parameters are $n = 0.99$, $q = 0.01$ and $\theta = 0.25$. The dashed straight lines correspond to the power laws predicted in the text for the asymptotic and corresponding intermediate asymptotic regimes.

The intermediate power law decay laws with exponents $1 - (m+1)\theta$ hold only when this exponent is positive, i.e., for $m < \frac{1}{\theta} - 1$. To understand what happens for larger m's, a more careful analysis is required, which is presented in Appendix B, which shows that formula (92) still holds and predicts that $g_m(t)$ is an increasing function of time for times $\alpha t < (\gamma\beta)^{1/\theta}$ before decreasing again with the standard asymptotic power law $\sim 1/t^{1+\theta}$. This is summarized by Figure 10, which plots the time-dependence of the rates $g_m(t) := \rho^{k,k+m}(t)$ of events of type $s = k + m$ triggered by a mother of type k for $m = 0$ to 5, with $\theta = 1/3$. One can clearly observe the existence of the intermediate power asymptotics (Eq. (B.24) in Appendix B and expression (92)) for different values of m. When inequality (B.25) of Appendix B holds, the intermediate asymptotics are not decaying but growing as a function of time, as predicted by expression (B.24) of Appendix B and (92).

6 One-dimensional chain of nearest-neighbor-type triggering

6.1 Definitions

A natural extension to the above one-dimensional chain of directed triggering discussed in the previous section includes feedbacks from events of type $k + 1$ to type k. The example treated in the present section corresponds to fully symmetry mutual excitations confined to nearest neighbor in the sense of event types: $k \leftrightarrow k + 1$. Mathematically, this is described by a symmetric circulant matrix $\hat{N}$ of the average numbers $n_{k,s}$ of first-generation events of different types triggered by a mother of a fixed type.

We assume that all diagonal elements are equal to some constant χ (same self-triggering abilities) and all their
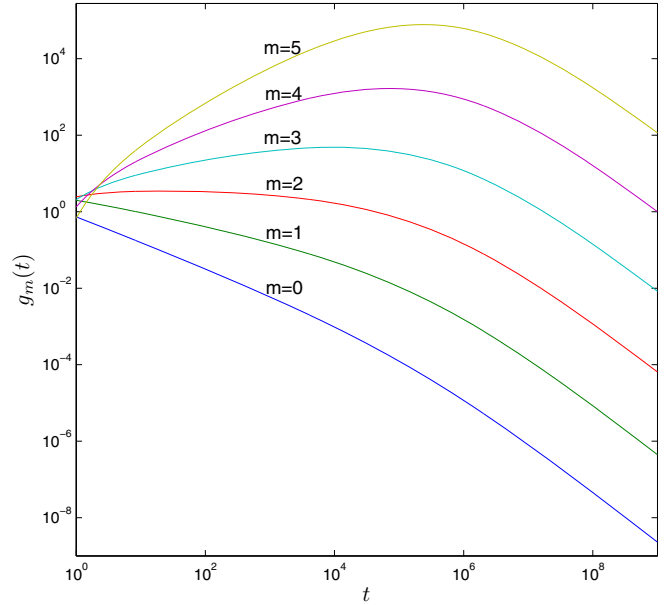

Fig. 10. Time-dependence of the rates $g_m(t) := \rho^{k,k+m}(t)$ of events of type $s = k + m$ triggered by a mother of type k, for $m = 0$ (same type of events as the initial triggering mother) and $m = 1, 2, 3, 4$ and 5. Here, $\theta = 1/3$ and $\gamma := \ln\left(\frac{1+q}{n}\right) = 0.01$. The number of summands used in the sum (B.7) of Appendix B is $N = 1500$.

neighbors are equal to some different constant ξ (same mutual triggering abilities). The elements $n_{1,m}$ and $n_{m,1}$ are also equal to ξ to close the chain of mutual excitations between events of type 1 and of type m. Restricting to $m = 6$ for illustration purpose, the corresponding matrix $\hat{N}$ reads

$$\hat{N} = \begin{bmatrix} \chi & \xi & 0 & 0 & 0 & \xi \\ \xi & \chi & \xi & 0 & 0 & 0 \\ 0 & \xi & \chi & \xi & 0 & 0 \\ 0 & 0 & \xi & \chi & \xi & 0 \\ 0 & 0 & 0 & \xi & \chi & \xi \\ \xi & 0 & 0 & 0 & \xi & \chi \end{bmatrix} \qquad (93)$$

where

$$\chi = \frac{n}{1+q}, \qquad \xi = \frac{nq}{2(1+q)} \quad \Rightarrow \quad \chi + 2\xi = n. \quad (94)$$

As before, the parameter q quantifies the "strength" of the interactions between events of different types. Here, n represents the total number of first-generation events of all types that are generated by a given mother of fixed arbitrary type. Figure 11 provides the geometrical sense of matrix $\hat{N}$ (93) for $m = 6$, where the circles represent the six types of events and the arrows denote their mutual excitation influences.

6.2 Analysis of the event rates $\rho^{k,s}(t)$

The Laplace transform $\tilde{\rho}^{k,k}(u)$ of the event rates $\rho^{k,k}(t)$ (30) of type k that are triggered by a mother of the same type k reads

$$\tilde{\rho}^{k,k}(u) := \tilde{\rho}(u) = A[n\tilde{f}(u), q], \qquad (95)$$
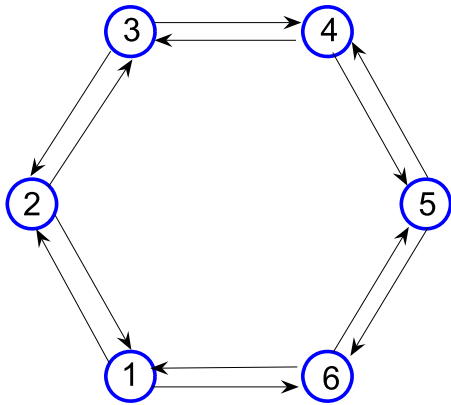
Fig. 11. Geometric sense of the matrix $\hat{N}$ for a one-dimensional chain of nearest-neighbor triggering in the space of event types.

where [73]

$$A(n,q) =$$
$$\frac{4n(1-n+q)^3 - (1-n+q)(5n-2q-2)n^2q^2 - n^4q^4}{4(1-n+q)^4 - 5(1-n+q)^2n^2q^2 + n^4q^4}. \tag{96}$$

Analogously, the Laplace transforms of the cross-rates $\tilde{\rho}^{k,s}(u)$ are

$$\tilde{g}_1(u) = B\left[n\tilde{f}(u), q\right],$$
$$\tilde{g}_2(u) = C\left[n\tilde{f}(u), q\right],$$
$$\tilde{g}_3(u) = D\left[n\tilde{f}(u), q\right], \tag{97}$$

where [73]

$$B(n,q) = \frac{nq(1+q)(2(1-n+q)^2 - n^2q^2)}{4(1-n+q)^4 - 5(1-n+q)^2n^2q^2 + n^4q^4},$$

$$C(n,q) = \frac{n^2q^2(1+q)(1-n+q)}{4(1-n+q)^4 - 5(1-n+q)^2n^2q^2 + n^4q^4},$$

$$D(n,q) = \frac{n^3q^3(1+q)}{4(1-n+q)^4 - 5(1-n+q)^2n^2q^2 + n^4q^4}. \tag{98}$$

Figure 12 shows the time-dependence of $\rho(t)$, $g_1(t)$, $g_2(t)$ and $g_3(t)$ for the case where the pdf $f(t)$ is a power law (48), for the parameters $n = 0.995$, $q = 0.01$, $\theta = 0.2$. One can observe a common power law asymptotic $\sim t^{-1-\theta} = t^{-1.2}$ at large times, as well as intermediate asymptotic power laws

$$\rho(t) \sim t^{-1+\theta} = t^{-0.8}, \quad g_1(t) \sim t^{-1+2\theta} = t^{-0.6},$$
$$g_2(t) \sim t^{-1+3\theta} = t^{-0.4}, \quad g_3(t) \sim t^{-1+4\theta} = t^{-0.2}. \tag{99}$$

7 Concluding remarks

We have presented a preliminary analysis of some temporal properties of multivariate self-excited Hawkes con-
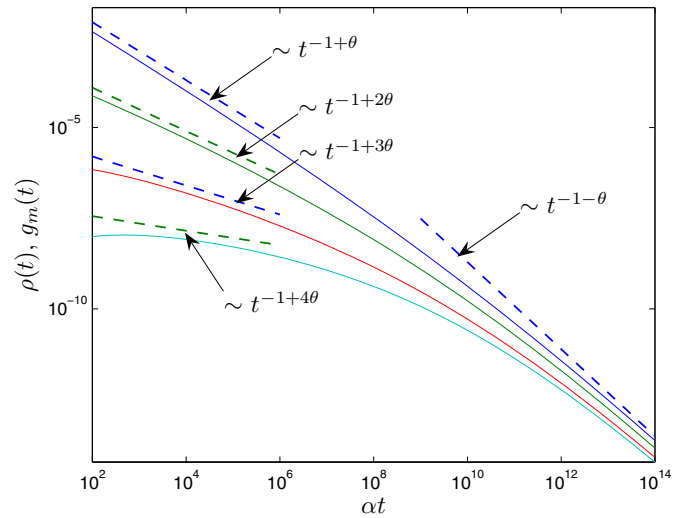

Fig. 12. Top to bottom: the solid lines represent the time-dependence of $\rho(t)$, $g_1(t)$, $g_2(t)$ and $g_3(t)$, in the case a one-dimensional chain of nearest-neighbor-type triggering in the space of types, with six types. The parameters are $n = 0.995$, $q = 0.01$, $\theta = 0.2$. The dashed straight lines show the asymptotic and intermediate asymptotic power laws predicted in the text.

ditional Poisson processes. These processes are very interesting candidates to model a large variety of systems with bursty events, for which past activity triggers future activity. The term "multivariate" refers to the property that events come in different types, with possibly different intra- and inter-triggering abilities. The richness of the generated time dynamics comes from the cascades of intermediate events of possibly different kinds, unfolding via random changes of types along the genealogy. We have developed the general formalism of the multivariate moment generating function for the cumulative number of first-generation and of all generation events triggered by a given mother event as a function of the current time t. We have obtained the general relations for the mean numbers of events triggered over all generations by a given event as a function of time. We have applied this technical and mathematical toolbox to several systems, characterized by different specifications on how events of a given type may trigger events of different types. In particular, for systems in which triggering between events of different types proceeds through a one-dimension directed or symmetric chain of influence in type space, we have discovered a novel hierarchy of intermediate asymptotic power law decays of the rate of triggered events as a function of the distance of the events to the initial shock in the space of types. We have been able to derive the time-dependence of the rates of events triggered from a given shock for distributions of waiting times of first-generation events that have either exponential or power law tails, for a variety of systems. Future directions of investigations include the study of more realistic networks in type-space and of the full distribution of even rates, beyond the mean dynamics reported here.

Appendix A: Proof of theorem 3.1

Substituting relation (13) in expression (14) leads to

$$
A_1^k(y_1, y_2, \ldots, y_m; t) =
$$
$$
\sum_{d_1=0}^{\infty} \cdots \sum_{d_m=0}^{\infty} \sum_{r_1=d_1}^{\infty} \cdots \sum_{r_m=d_m}^{\infty} P_k(r_1, \ldots, r_m)
$$
$$
\times \prod_{s=1}^{m} \binom{r_s}{d_s} [\mu_{k,s}(t) y_s]^{d_s} [1 - \mu_{k,s}(t)]^{r_s - d_s}. \quad (A.1)
$$

Inverting the order of the summations

$$
\sum_{d_s=0}^{\infty} \sum_{r_s=d_s}^{\infty} (\cdots) = \sum_{r_s=0}^{\infty} \sum_{d_s=0}^{r_s}, \quad (A.2)
$$

we rewrite expression (A.1) as

$$
A_1^k(y_1, y_2, \ldots, y_m; t) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_m=0}^{\infty} P_k(r_1, \ldots, r_m)
$$
$$
\times \sum_{d_1=0}^{r_1} \cdots \sum_{d_m=0}^{r_m} \prod_{s=1}^{m} \binom{r_s}{d_s}
$$
$$
\times [\mu_{k,s}(t) y_s]^{d_s} [1 - \mu_{k,s}(t)]^{r_s - d_s}, \quad (A.3)
$$

or equivalently

$$
A_1^k(y_1, y_2, \ldots, y_m; t) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_m=0}^{\infty} P_k(r_1, \ldots, r_m)
$$
$$
\times \prod_{s=1}^{m} \sum_{d_s=0}^{r_s} \binom{r_s}{d_s}
$$
$$
\times [\mu_{k,s}(t) y_s]^{d_s} [1 - \mu_{k,s}(t)]^{r_s - d_s}. \quad (A.4)
$$

Using the binomial formula

$$
\sum_{d_s=0}^{r_s} \binom{r_s}{d_s} [\mu_{k,s}(t) y_s]^{d_s} [1 - \mu_{k,s}(t)]^{r_s - d_s} = [1 + \mu_{k,s}(y-1)]^{r_s}, \quad (A.5)
$$

we obtain

$$
A_1^k(y_1, y_2, \ldots, y_m; t) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_m=0}^{\infty} P_k(r_1, \ldots, r_m)
$$
$$
\times \prod_{s=1}^{m} [1 + \mu_{k,s}(t)(y-1)]^{r_s}. \quad (A.6)
$$

In view of definition (8) of the MGF $A_1^k(y_1, y_2, \ldots, y_m)$, this last expression means that

$$
A_1^k(y_1, y_2, \ldots, y_m; t) = A_1^k [1 + \mu_{k,1}(t)(y_1 - 1), \ldots, 1
$$
$$
+ \mu_{k,m}(t)(y_m - 1)]. \quad (A.7)
$$

Using definition (16) of the function Q_k, we obtain relation (15).

Appendix B: Analysis of the behavior of $g_m(t) := \rho^{k,s}(t)$ for a one-dimensional chain of directed triggering of Section 5 for $n \to 1 + q$ when $1 - (m+1)\theta < 0$

Let us start with expression $\tilde{g}_m(u)$ (84) that we rewrite, omitting the nonessential factor ξ^m, as

$$
\tilde{g}_m(u) = \frac{\tilde{f}^m(u)}{[1 - \chi \cdot \tilde{f}(u)]^{m+1}}, \quad \text{with } \chi = \frac{n}{1+q}. \quad (B.1)
$$

Using the binomial formula

$$
\frac{1}{(1-x)^{m+1}} = \sum_{k=0}^{\infty} \binom{m+k}{k} x^k, \quad (B.2)
$$

expression (B.1) becomes

$$
\tilde{g}_m(u) = \sum_{k=0}^{\infty} \binom{m+k}{k} \chi^k \tilde{f}^{m+k}(u). \quad (B.3)
$$

As we are interested in the case where the pdf $f(t)$ has the power asymptotic $f(t) \sim 1/t^{1+\theta}$, with $0 < \theta < 1$, it is convenient to use for $f(t)$ one special representative of the functions presenting this asymptotic power law behavior, namely the one-sided Lévy stable distribution of order θ [74], that we refer to as $f_\theta(t)$. Its Laplace transform is

$$
\tilde{f}_\theta(u) = e^{-u^\theta}, \qquad 0 < \theta < 1. \quad (B.4)
$$

Accordingly, relation (B.3) takes the form

$$
\tilde{g}_m(u) = \sum_{k=0}^{\infty} \binom{m+k}{k} \chi^k e^{-(m+k)u^\theta}. \quad (B.5)
$$

Taking the inverse Laplace transform of (B.5) provides us with the exact expression

$$
g_m(t) = \sum_{k=0}^{\infty} \binom{m+k}{k} \frac{\chi^k}{(m+k)^{1/\theta}} f_\theta \left(\frac{t}{(m+k)^{1/\theta}} \right). \quad (B.6)
$$

In order to analyze (B.6), it is convenient to rewrite it as

$$
g_m(t) = \frac{\theta}{t^{1+\theta}} \sum_{k=0}^{\infty} S_m(k+m) Q_\theta \left(\frac{m+k}{t^\theta} \right) e^{-\gamma k}, \quad (B.7)
$$

where

$$
\gamma = \ln \left(\frac{1}{\chi} \right) = \ln \left(\frac{1+q}{n} \right) > 0 \text{ for } n < 1 + q. \quad (B.8)
$$

We have defined the functions

$$S_m(x) := x\binom{x}{m}, \qquad x \geqslant m, \qquad (B.9)$$

and

$$Q_\theta(x) := \frac{1}{\theta x^{1+1/\theta}} f_\theta\left(\frac{1}{x^{1/\theta}}\right), \qquad x \geqslant 0, \qquad (B.10)$$

such that

$$\frac{1}{x^{1/\theta}} f_\theta\left(\frac{t}{x^{1/\theta}}\right) = \frac{x\,\theta}{t^{\theta+1}} Q_\theta\left(\frac{x}{t^\theta}\right). \qquad (B.11)$$

In order to extract the relevant information from expression (B.7) for $g_m(t)$, we need to discuss some properties of the two functions $S_m(x)$ and $Q_\theta(x)$.

1. **Properties of the function $S_m(x)$.** The function $S_m(x)$ is a finite sum of power functions of the argument x

$$S_m(x) = \sum_{r=1}^{m} a_{r,m}\, x^{r+1}. \qquad (B.12)$$

In particular,

$$S_0(x) = x, \quad S_1(x) = x^2, \quad S_2(x) = \frac{1}{2}\,x^3 - \frac{1}{2}\,x^2,$$

$$S_3(x) = \frac{1}{6}\,x^4 - \frac{1}{2}\,x^3 + \frac{1}{3}\,x^2,$$

$$S_4(x) = \frac{1}{24}\,x^5 - \frac{1}{4}\,x^4 + \frac{11}{24}\,x^3 - \frac{1}{4}\,x^2. \qquad (B.13)$$

2. **Properties of the function $Q_\theta(x)$.** For $\theta \in (0, 1/2)$, this function is at least exponentially decaying with increasing x. Accordingly, its moments of any order $r > 0$ are finite and given by

$$M(r) := \int_0^\infty x^r\, Q_\theta(x)dx = \frac{\Gamma(r+1)}{\Gamma(r\theta+1)}. \qquad (B.14)$$

Moreover, the value of $Q_\theta(x)$ at $x = 0$ is equal to

$$Q_\theta(x = 0) = \frac{1}{\Gamma(1-\theta)}. \qquad (B.15)$$

For the particular cases $\theta = 1/2$ and $\theta = 1/3$, the function $Q_\theta(x)$ can be expressed in explicit form:

$$Q_{1/2}(x) = \frac{1}{\sqrt{\pi}} \exp\left(-\frac{x^2}{4}\right),$$

$$Q_{1/3}(x) = \sqrt[3]{9}\,\mathrm{Ai}\left(\frac{x}{\sqrt[3]{3}}\right). \qquad (B.16)$$

We study the behavior of $g_m(t)$ given by expression (B.7) for $\gamma m \ll 1, t^\theta \gg 1$ and $m\,t^{-\theta} \ll 1$. To leading order and without essential error, we may replace the discrete sum (B.7) by the continuous integral

$$g_m(t) \simeq \frac{\theta}{t^{1+\theta}} \int_m^\infty S_m(x)\, Q\left(\frac{x}{t^\theta}\right) e^{-\gamma x} dx. \qquad (B.17)$$

Using the following change of variable of integration

$$x \quad \mapsto \quad y = \frac{x}{t^\theta} \quad \leftrightarrow \quad x = t^\theta\, y, \qquad (B.18)$$

we obtain

$$g_m(t) \simeq \frac{\theta}{t} \int_0^\infty S_m(t^\theta y) Q_\theta(y) e^{-\gamma t^\theta y} dy. \qquad (B.19)$$

Using relation (B.12), this yields

$$g_m(t) \simeq \frac{\theta}{t} \sum_{r=1}^{m} a_{r,m} t^{(r+1)\theta} G_{r,m}(t), \qquad (B.20)$$

where

$$G_{r,m}(t) = \int_0^\infty y^{r+1} Q_\theta(y) e^{-\gamma t^\theta y} dy. \qquad (B.21)$$

The intermediate asymptotic regime corresponds to the time domain $\gamma t^\theta \ll 1$ with $t^\theta \gg 1$. In this case, the exponential in the integral (B.21) can be replaced by unity. Using relations (B.14), we obtain that

$$G_{r,m} = \int_0^\infty y^{r+1} Q_\theta(y) dy = \frac{\Gamma(r+2)}{\Gamma[(r+1)\theta+1]}, \quad \gamma t^\theta \ll 1, \qquad (B.22)$$

is time-independent. Accordingly, the mean rate $g_m(t)$ (B.20) is found as the sum of power law functions

$$g_m(t) \simeq \theta t^{(m+1)\theta-1} \sum_{r=1}^{m} a_{r,m} G_{r,m} t^{(r-m)\theta}. \qquad (B.23)$$

Taking into account that we consider the case $t^\theta \gg 1$ (that allowed us to use the integral approximation (B.17)), we obtain the sought power law intermediate asymptotic

$$g_m(t) \simeq \frac{\theta(m+1)}{\Gamma[(m+1)\theta+1]} t^{(m+1)\theta-1} \sim t^{(m+1)\theta-1},$$

$$1 \ll t \ll \gamma^{-1/\theta}. \qquad (B.24)$$

This recovers the result (92) presented in the main text. In addition, it makes more precise what happens for

$$(m+1)\theta > 1. \qquad (B.25)$$

In this case, $g_m(t)$ starts as a growing function of t up to $t \simeq \gamma^{-1/\theta}$. This retrieves the same qualitative behavior found when $f(t)$ is an exponential, which has been analyzed in Section 5.3 and represented in Figure 7.

Of course, at times $t \gg \gamma^{-1/\theta}$, this growth is replaced by the standard power law decay $\sim 1/t^{1+\theta}$. Indeed, for $\gamma t^\theta \gg 1$, the integral (B.21) is approximately equal to

$$G_{r,m}(t) \simeq Q_\theta(0) \int_0^\infty y^{r+1} e^{-\gamma t\theta y} dy$$

$$= \frac{(r+1)!}{\Gamma[1-\theta]} \gamma^{-r-2} t^{-(r+2)\theta}. \qquad (B.26)$$

Substituting this relation into (B.20) yields

$$g_m(t) \simeq \frac{\theta}{t^{1+\theta}} \sum_{r=1}^{m} a_{r,m} \frac{(r+1)!}{\Gamma[1-\theta]} \gamma^{-r-1} \sim \frac{1}{t^{1+\theta}}, \quad t \gg \gamma^{-1/\theta}.$$
(B.27)

Figure 10 in the text sums up these results by plotting the time-dependence of the rates $g_m(t) := \rho^{k,k+m}(t)$ of events of type $s = k + m$ triggered by a mother of type k for $m = 0$ to 5. One can clearly observe the existence of the intermediate power asymptotics (B.24) for different values of m. When inequality (B.25) holds, the intermediate asymptotics are not decaying but growing as a function of time, as predicted by expression (B.24).

References

1. A.G. Hawkes, J. Roy. Statist. Soc. Ser. B **33**, 438 (1971)
2. A.G. Hawkes, Spectra of some mutually exciting point processes with associated variables, in *Stochastic Point Processes*, edited by P.A.W. Lewis (Wiley, New York, 1972), pp. 261–271
3. A.G. Hawkes, L. Adamopoulos, Bull. Int. Stat. Inst. **45**, 454 (1973)
4. A.G. Hawkes, D. Oakes, J. Appl. Prob. **11**, 493 (1974)
5. P. Brémaud, *Point Processes and Queues* (Springer, New York, 1981)
6. R. Cont, P. Tankov, *Financial Modelling with Jump Processes* (CRC Press UK, London, 2004)
7. D.J. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer Series in Statistics (Springer, New York, 2007)
8. P. Horowitz, W. Hill, *The Art of Electronics*, 2nd edn. (Cambridge University Press, Cambridge, 1989), pp. 431–432
9. E.W. Montroll, H. Scher, J. Stat. Phys. **9**, 101 (1973)
10. H. Scher, E.W. Montroll, Phys. Rev. B **12**, 2455 (1975)
11. A. Helmstetter, D. Sornette, J. Geophys. Res. **107**, 2237 (2002)
12. Y. Ogata, J. Am. Stat. Assoc. **83**, 9 (1988)
13. Y. Ogata, Tectonophysics **169**, 159 (1989)
14. Y. Ogata, Ann. Inst. Stat. Mech. **50**, 379 (1998)
15. Y. Ogata, Pure Appl. Geophys. **155**, 471 (1999)
16. A. Saichev, D. Sornette J. Geophys. Res. **112**, B04313 (2007)
17. D. Sornette, I. Osorio, Prediction, in *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Physics and Engineering*, edited by I. Osorio, H.P. Zaveri, M.G. Frei, S. Arthurs (CRC Press, 2010), arXiv:1007.2420
18. V. Chavez-Demoulin, A.C. Davison, A.J. McNeil, Quant. Finance **5**, 227 (2005)
19. L. Bauwens, N. Hautsch, Modelling Financial High Frequency Data Using Point Processes, *Handbook of Financial Time Series*, Part 6, pp. 953–979, DOI: 10.1007/978-3-540-71297-8_41, 2009
20. E. Errais, K. Giesecke, L.R. Goldberg, Affine Point Processes and Portfolio Credit Risk (2010). Available at SSRN: http://ssrn.com/abstract=908045
21. S. Azizpour, K. Giesecke, G. Schwenkler, Exploring the Sources of Default Clustering, working paper, Stanford University, 2012, http:www.stanford.edu/dept/MSandE/cgi-bin/people/faculty/giesecke/pdfs/exploring.pdf
22. Y. Aït-Sahalia, J. Cacho-Diaz, R.J.A. Laeven, Modeling financial contagion using mutually exciting jump processes, working paper 15850, 2013, http://www.princeton.edu/∼yacine/contagion.pdf
23. M. Salmon, W.W. Tham, Preferred Habitat, Time Deformation and the Yield Curve, working paper (Journal of Financial Markets), 2008
24. C. Scholz, *The Mechanics of Earthquakes and Faulting*, 2nd edn. (Cambridge University Press, Cambridge, 2003)
25. Y. Malevergne, D. Sornette, *Extreme Financial Risks (From Dependence to Risk Management)* (Springer, Heidelberg, 2006)
26. M. Gomez Rodriguez, J. Leskovec, A. Krause, in *Proceedings of the 16th ACM SIGKDD* (2010)
27. J. Leskovec, D. Huttenlocher, J. Kleinberg, in *Proceedings of the 19th Int. Conf. WWW* (2010)
28. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Political Economy of Institutions and Decisions) (Cambridge University Press, Cambridge, 1990)
29. K. Crowston, J. Howison, K. Crowston, J. Howison, Knowl. Technol. Policy **18**, 65 (2006)
30. G. von Krogh, S. Spaeth, K.R. Lakhani, Res. Policy **32**, 1217 (2003)
31. S. Krishnamurthy, Technol. Policy **18**, 17 (2006)
32. C.Y. Baldwin, K.B. Clark, Manage. Sci. **52**, 1116 (2006)
33. A. MacCormack, J. Rusnak, C.Y. Baldwin, Manage. Sci. **52**, 1015 (2006)
34. D. Challet, A. Lombardoni, Phys. Rev. E **70**, 046109 (2004)
35. T. Maillart, D. Sornette, S. Spaeth, G. von Krogh, Phys. Rev. Lett. **101**, 218701 (2008)
36. A.-L. Barabási, Nature **435**, 207 (2005)
37. T. Maillart, D. Sornette, S. Frei, T. Duebendorfer, A. Saichev, Phys. Rev. E **83**, 056101 (2011)
38. A. Saichev, D. Sornette, Phys. Rev. E **81**, 016108 (2009)
39. J. Zhuang, D. Vere-Jones, H. Guan, Y. Ogata, L. Ma, Pure Appl. Geophys. **162**, 1367 (2006)
40. Y.Y. Kagan, L. Knopoff, J. Geophys. Res. **86**, 2853 (1981)
41. A. Helmstetter, Phys. Res. Lett. **91**, 058501 (2003)
42. D. Sornette, M.J. Werner, J. Geophys. Res. **110**, B08304 (2005)
43. D. Marsan, O. Lengliné, Science **319**, 1076 (2008)
44. Y. Ogata, IEEE Trans. Inf. Theory **IT-27**, 23 (1981)
45. Y. Ogata, J. Phys. Earth **31**, 115 (1983)
46. Y. Ogata, H. Akaike, J. Roy. Stat. Soc. Ser. B **44**, 102 (1982)
47. Y. Ogata, J. Zhuang, Tectonophysics **413**, 13 (2006)
48. T. Ozaki, Ann. Inst. Statist. Math. **31**, 145 (1979)
49. D. Sornette, S. Utkin, Phys. Rev. E **79**, 061110 (2009)
50. J. Zhuang, Ph.D. thesis, The Graduate University for Advanced Studies, 2002
51. J. Zhuang, Y. Ogata, D. Vere-Jones, J. Am. Statist. Assoc. **97**, 369 (2002)
52. J. Zhuang, Y. Ogata, D. Vere-Jones, J. Geophys. Res. **109**, B05301 (2004)
53. A. Helmstetter, D. Sornette, Phys. Rev. E **66**, 061104 (2002)
54. A. Helmstetter, D. Sornette, Geophys. Res. Lett. **30**, 2069 (2003), DOI:10.1029/2003GL017670
55. A. Helmstetter, D. Sornette, J. Geophys. Res. **108**, 2482 (2003)

56. A. Helmstetter, D. Sornette, J.-R. Grasso, J. Geophys. Res. **108**, 2046 (2003)
57. A. Saichev, A. Helmstetter, D. Sornette, Pure Appl. Geophys. **162**, 1113 (2005)
58. A. Saichev, D. Sornette, Phys. Rev. E **70**, 046123 (2004)
59. A. Saichev, D. Sornette, Phys. Rev. E **71**, 056127 (2005)
60. A. Saichev, D. Sornette, Eur. Phys. J. B **51**, 443 (2006)
61. D. Sornette, A. Helmstetter, Phys. Rev. Lett. **89**, 158501 (2002)
62. D. Sornette, S. Utkin, A. Saichev, Phys. Rev. E **77**, 066109 (2008)
63. T.E. Harris, *The Theory of Branching Processes* (Springer, Berlin, 1963)
64. D. Sornette, Proc. Natl. Acad. Sci. USA **99**, 2522 (2002)
65. A. Saichev, D. Sornette, Eur. Phys. J. B **75**, 343 (2010)
66. T.J. Liniger, *Multivariate Hawkes Processes*, Ph.D. Diss. ETH No. 18403, ETH Zurich, 2009
67. R. Crane, D. Sornette, Proc. Natl. Acad. Sci. USA **105**, 15649 (2008)
68. F. Deschatres, D. Sornette, Phys. Rev. E **72**, 016112 (2005)
69. D. Sornette, Endogenous versus exogenous origins of crises, in the monograph entitled "Extreme Events in Nature and Society", Series: *The Frontiers Collection*, edited by S. Albeverio, V. Jentsch, H. Kantz (Springer, Heidelberg, 2005), pp. 95–119, `arXiv:physics/0412026`
70. D. Sornette, F. Deschatres, T. Gilbert, Y. Ageon, Phys. Rev. Lett. **93**, 228701 (2004)
71. A. Sornette, D. Sornette, Geophys. Res. Lett. **26**, 1981 (1999)
72. D. Sornette, A. Helmstetter, Physica A **318**, 577 (2003)
73. A. Saichev, D. Sornette, Eur. Phys. J. B **83**, 271 (2011)
74. V.M. Zolotarev, One-dimensional Stable Distributions, *Translations of Mathematical Monographs* (Am. Math. Soc., Providence, 1986), Vol. 65