Trend detection in social networks using Hawkes processes

Julio Cesar Louzada Pinto
Institut Mines-Telecom
Telecom SudParis, RST Department
UMR CNRS 5157, France
Email: julio.louzada_pinto@telecom-sudparis.eu

Tijani Chahed
Institut Mines-Telecom
Telecom SudParis, RST Department
UMR CNRS 5157, France
Email: tijani.chahed@telecom-sudparis.eu

Eitan Altman
INRIA Sophia Antipolis Mediterranée
06902 Sophia-Antipolis Cedex, France
Email: eitan.altman@inria.fr

***Abstract*—We develop in this paper a trend detection algorithm, designed to find trendy topics being disseminated in a social network. We assume that the broadcasts of messages in the social network is governed by a self-exciting point process, namely a Hawkes process, which takes into consideration the real broadcasting times of messages and the interaction between users and topics. We formally define trendiness and derive trend indices for each topic being disseminated in the social network. These indices take into consideration the time between the detection and the message broadcasts, the distance between the real broadcast intensity and the maximum expected broadcast intensity, and the social network topology. The proposed trend detection algorithm is simple and uses stochastic control techniques in order calculate the trend indices. It is also fast and aggregates all the information of the broadcasts into a simple one-dimensional process, thus reducing its complexity and the quantity of necessary data to the detection.**

I. Introduction

This paper introduces a novel trend detection algorithm which seeks to discover trendy topics being disseminated in a social network.

Since we are dealing with social networks, we cannot use classical trend detection algorithms [1], [2], as they simply use text mining and queuing techniques and do not grasp the full relationship between users and contents in the social network. This idea of leveraging social and textual contents is quite recent, with works as [3], [4] shedding some light into the matter.

Thus, in order to fully exploit the social ties between users and information in social networks, the proposed algorithm bases itself on information diffusion models [5], [6], [7], [8], or more specifically on a Hawkes-based model for information diffusion in social networks [9], [10], [11], [12]. Being aware that the adoption of a parametric Hawkes model for information diffusion could theoretically restrict the usefulness of the trend detection algorithm, we are persuaded that it remains interesting for several reasons: 1) it allows leveraging on the knowledge of the influences between users and contents, 2) it allows fully exploring the real time of broadcasts, 3) it allows leveraging on the knowledge of users intrinsic (or exogenous) rates, 4) its intensity represents the propensity of users to broadcasts topics at each time, thus serving as proxy for the activity level of topics and users in the social network [13], etc.

Such information diffusion frameworks were already adopted under different scenarios, as did Kempe *et al.* in their seminal paper [6] by developing a framework based on submodular functions to detect the optimal seed group in order to diffuse a content, using the so-called independent cascade propagation model [5]. Other examples of such methodology are [8], where Altshuler *et al.* derive a method capable of predicting future trends based on the analysis of past social interactions between community members of a scale-free network; [14], where Cheng *et al.* propose a framework for addressing cascade prediction problems, motivated by a view of cascades as complex dynamic objects passing through successive stages while growing; and [15], where Leskovec *et al.* develop a scalable framework for tracking short, distinctive phrases (so-called memes [16]) that travel relatively intact through online text, providing a representation of the news cycle.

As already mentioned, we adopt in the present work an information diffusion approach and use specifically point processes, namely Hawkes processes [17], [18], to track the exact diffusion times of the information cascades in a social network, taking into account the interaction between the topics, the users and the underlying social network structure.

We assume that there exist different topics being disseminated in a social network and we employ the Hawkes process to count the number of broadcasts of these topics by each of the users in the social network. We say that a topic is *trendy* if it has a rapid increase in its broadcasting Hawkes intensity. These topic intensities are combinations of the users broadcasting intensities, where each user contributes to the topic intensities with a measure of his impact on the network, proportional to his network outgoing eigenvector centrality [19].

A trendy topic has then a burst in its broadcasting in the network, which corresponds to an increase in its broadcasting intensity, or a *peak*. Our algorithm thus seeks the "peaks" in the intensity of the underlying Hawkes process in order to determine the topics that are likely to be trendy in the future.

In order to search these "peaks" in the Hawkes intensity, we use scaling techniques [20] on the Hawkes intensity to transform it into a Brownian diffusion, which allows the use of the well known machinery of stochastic control [21] to implement our trend detection algorithm.

Contributions

The contributions of this paper are the following:

- *To the best of our knowledge, this is the first trend detection algorithm that uses point processes and stochastic control techniques*. These techniques are successfully used in many other fields, and are complementary tools to machine learning and text mining techniques, hence providing more diversified treatments for this kind of problem.
- *The difference between the proposed trend detection algorithm and one that looks solely at the topics with the largest number of broadcasts is that we aim to detect those topics that are trendy but do not necessarily with large number of broadcasts*. Indeed, the most straightforward approach would be to look at the point process intensities and choose those topics with the highest intensities. Our approach is different: we do not compare topics between themselves, but rather compare the topic intensities with their maximum expected intensities, meaning that topics that do not have yet large intensities can indeed become trendy. Still, our algorithm is also able to capture the trendiness coming from large intensities.

The remainder of this paper is organized as follows. In section II, we present the adopted model of information diffusion in the social network using Hawkes processes. In section III, we define trendiness in our context, detail our trend detection algorithm and derive the trend indices for topics of messages broadcasted in the social network. In section IV, we illustrate our algorithm using two different datasets. Section V eventually concludes the paper.

II. Information diffusion

We start the theoretical study of our trend detection algorithm by adopting a model for information diffusion in social networks. This model is based on point processes, or more precisely on the so-called linear Hawkes process [17], [18].

A. A Hawkes model

Hawkes-based information diffusion models are widely adopted to model information diffusion in social networks [9], [10], [11], [12]. This is due to several reasons, which are nonexhaustively listed here:

- They are point processes [22], and as such they are designed to model discrete events in networks such as posting, sharing, tweeting, liking, digging, etc.
- Hawkes processes are self-excited processes, i.e., the probability of a future event increases with the occurrence of past events.
- They possess a simple and linear structure for their intensity (the conditional expectation of an occurrence of an event, at each time).
- They present simple maximum likelihood formulas [22], [23], which facilitates a maximum likelihood estimation of the parameters.
- A linear Hawkes process can be seen as a Poisson cluster process [24], which permits the distinction of two regimes: a stationary (or stable) regime in which the intensity processes has a stationary and nonexplosive version, and a nonstationary (or unstable) regime, in which the process has an unbounded number of events (see [17], [25] for details).
- It easily allows extensions from the basic model, such as multiple social networks [26], dynamic/temporal networks [27], seasonality and/or time-dependence for the intrinsic diffusion rate of users [12], etc.

Thus, after listing the properties of Hawkes processes that are interesting when modeling information diffusion in social networks, we start the detailed description of the adopted information diffusion model in this paper:

We represent our social network as a communication graph $G = (V, E)$, where V is the set of users with cardinality $\sharp V = N$ and E is the edge set, i.e., the set with all the possible communication links between users, as in [11]. We assume this graph to be directed and weighted, and coded by an inward adjacency matrix J such that $J_{i,j} > 0$ if user j is able to broadcast messages to user i, or $J_{i,j} = 0$ otherwise. If one thinks about Twitter, $J_{i,j} > 0$ means that user i follows user j and receives the news published by user j in his or her timeline.

We assume that users in this social network broadcast messages (post, share, comment, tweet, retweet, etc.) during a time interval $[0, \tau]$. These messages represent information about K predefined[1] topics (economics, religion, culture, politics, sports, music, etc.), and at each event the broadcasted message concerns one and only one specific topic among these K different ones.

When broadcasting, users may influence others to broadcast. For example: when tweeting, the user's followers may find the tweet interesting and retweet it to their friends and followers, generating then a cascade of tweets.

We assume that these influences are divided into two categories: user-user influences and topic-topic influences. For example, during these retweeting cascade, users may react differently to the content of the tweet in question, which of course may imply a different influence of this particular tweet among users. By the same token, the followers in question may respond differently depending on the broadcaster, since people influence others differently in social networks.

The influences are coded by the $N \times N$ matrix J and the $K \times K$ matrix B, such that $J_{i,j} \geq 0$ is the (possible) influence of user i over user j and $B_{c,k} \geq 0$ is the (possible) influence of topic c over topic k.

In light of this explanation, we assume that the cumulative number of messages broadcasted by users is a linear Hawkes process X, where $X_t^{i,k}$ represents the cumulative number of messages of topic k broadcasted by user i until time $t \in [0, \tau]$.

Let $\mathcal{F}_t = \sigma(X_s,\ s \leq t)$ be the filtration generated by the Hawkes process X. Our Hawkes process is then a $\mathbb{R}^{N \times K}$

[1]In our work, we rely on text mining techniques only to classify the broadcasted messages into different topics.

point process with intensity $\lambda_t = \lim_{\delta \searrow 0} \mathbb{E}[X_{t+\delta} - X_t|\mathcal{F}_t]/\delta$ defined as

$$\lambda_t^{i,k} = \mu^{i,k} + \sum_j \sum_c J_{i,j} B_{c,k} \int_0^{t-} \phi(t-s) dX_s^{j,c},$$

where $\mu^{i,k} \geq 0$ is the intrinsic (or exogenous) intensity of the user i for broadcasting messages of topic k and $\phi(t)$ is a nonnegative causal kernel responsible for the temporal impact of the past interactions between users and topics, satisfying $||\phi||_1 = \int_0^\infty \phi(u)du < \infty$.

Remark: Two common time-decaying functions are $\phi(t) = e^{-\omega t}.\mathbb{I}_{\{t>0\}}$ a light-tailed exponential kernel [11] and $\phi(t) = (a+t)^{-b}.\mathbb{I}_{\{t>0\}}$ a heavy-tailed power-law kernel [9].

The intensity can be seen in matrix form as

$$\lambda_t = \mu + J(\phi * dX)_t B, \tag{1}$$

where $(\phi * dX)_t$ is the $N \times K$ convolution matrix defined as $(\phi * dX)_t^{i,k} = \int_0^t \phi(t-s) dX_s^{i,k}$.

Remark: This paper is not concerned with the estimation of the Hawkes parameters μ, J and B. Maximum likelihood estimation and L^2 contrast minimization procedures can be used to estimate J and B, as in [11], [12], [28].

B. Stationary regime

As already mentioned in subsection II-A, one of the main properties of linear Hawkes processes is that they have a narrow link with branching processes with immigration [24], which gives us the following result (whose proof is well explained in [17], [25]):

Lemma 1. *We have that the linear Hawkes process X_t admits a version with stationary increments if and only if it satisfies the following stability condition*[2]

$$sp(J)sp(B)||\phi||_1 < 1. \tag{2}$$

III. Discovering trendy topics

After defining in details the adopted information diffusion framework serving as foundation for our trend detection algorithm, we continue towards the real goal of this paper: *to derive a Hawkes-based trend detection algorithm.*

The proposed algorithm takes into consideration the entire history of the Hawkes process X_t for $t \in [0,\tau]$ and makes a prediction for the trendiest topics at time τ, based on trend indices $\mathcal{I}^k$, $k \in \{1,2,\cdots,K\}$. It consists on the following steps:

1) Perform a temporal rescaling of the intensity following the theory of nearly unstable Hawkes processes [20], which gives a Cox-Ingersoll-Ross (CIR) process [29] as the limiting rescaled process.
2) Search the expected maxima of the rescaled intensities for each topic $k \in \{1,2,\cdots,K\}$, with the aid of the limit CIR process. This task is achieved by solving stochastic control problems V_k following the theory developed in [21], which measure the deviation of the rescaled intensities with respect to their stationary mean.
3) Generate from the control problems V_k time-dependent indices $\mathcal{I}_t^k$, which measure the peaks of each topic during the whole dissemination period $[0,\tau]$. We create then the trend indices $\mathcal{I}^k = \int_0^\tau \mathcal{I}_t^k dt$ for each topic $k \in \{1,2,\cdots,K\}$.

[2]Where for a squared matrix A we denote by $sp(A)$ its spectral radius, i.e., $sp(A) = \sup\{|\lambda| \mid det(A - \lambda\mathbb{I}) = 0\}$.

A. Trendy topics and rescaling

As our algorithm is based on the assumption that *a trendy topic is one that has a rapid and significant increase in the number of broadcasts*, a major tool in the development of this trend detection algorithm is the rescaling of nearly unstable Hawkes processes, developed by Jaisson and Rosenbaum in [20].

As already mentioned in section II, Hawkes processes possess two distinct regimes: a *stable* regime, where the intensity λ_t possesses a stationary version and the number of broadcasts is bounded almost surely, and an *unstable* regime where the number of broadcasts is unbounded almost surely.

The intuition behind the rescaling is the following: since we want to measure topics that have a burst in the number of broadcasted messages, we place ourselves between the stable and unstable regime, where the stability equation (2) is barely satisfied, i.e., $sp(J)sp(B)||\phi||_1 \sim 1$ - a Hawkes process satisfying this property is called *nearly unstable* [20]. By placing ourselves in the stable regime, the Hawkes process still possesses a limited number of broadcasted messages, but as we approach the unstable regime, the number of broadcasted messages increases (which could represent trendiness). Our trend detection algorithm uses hence this rationale in order to transform the Hawkes intensity λ into a Brownian diffusion, for which stochastic control techniques exist and are easy to implement.

The rescaling works thus in the following fashion: as the trendy data has a large number of broadcasts, we artificially "push" the Hawkes process X to the unstable regime when estimating the parameters μ, B, J and ϕ, in order to accommodate this large quantity of broadcasts. Then, by rescaling the intensity λ, it converges in law when $\tau \to \infty$ to a one-dimensional Cox-Ingersoll-Ross (CIR) process (see theorem 1), whose deviation to the stationary mean is studied using stochastic control techniques, or more precisely, by detecting its expected maxima [21].

Remark: In order to find the most appropriate nearly unstable regime for the Hawkes process X, the choice of the time horizon τ is crucial, as it determines the timescale of the predicted trends. It means that if ones uses τ measured in seconds, the prediction regards what happens in the seconds after the prediction period $[0,\tau]$, if one uses τ measured in days, the prediction regards what happens in the next day or days after the prediction period $[0,\tau]$, etc.

B. Topic trendiness

We recall the definition of *trendiness* in our context of information diffusion: *a trendy topic is one that has a rapid and significant increase in the number of broadcasts.*

Although this idea is fairly simple, care must be taken: the definition must take into consideration the social network in question, since users do not affect it on the same way. For example: if Barack Obama tweets about climate change, one may assume that climate change may become a trendy topic, but if an anonymous user tweets about the same topic, one has less argument to believe that the topic will become trendy. By the same token, if a group composed of many people start tweeting about the latest iPhone, one may consider it a trendy topic, but if only a small group of friends starts tweeting about it, again, one may not be inclined to think so.

Let us discuss it in more details: since the intensity λ_t is associated with the expected increase in broadcasts at time t, we use λ as base measure for the trendiness. Moreover, by the previous paragraph, we must also weight the intensity λ with a user-network measure responsible for the impact of users on the network. In our case, this user-network measure is the outgoing network eigenvector centrality of users [19].

Mathematically speaking, let v^T be the left-eigenvector of the user-user interaction matrix J, related to the leading[3] eigenvalue $\nu > 0$. Since v is the leading eigenvector of J^T - the outward weighted adjacency matrix of the communication graph in our social network - it represents the outgoing centrality of the network (also known as eigenvector centrality, similar to the pagerank algorithm [19], [30]) and consequently the users' impact on the network, as desired.

Multiplying Eqn. (1) in the left by v^T we have that

$$\begin{aligned} v^T\lambda_t &= v^T\mu + v^T J(\phi * dX)_t B \\ &= v^T\mu + \nu v^T(\phi * dX)_t B \\ &= v^T\mu + \nu(\phi * v^T dX)_t B. \end{aligned}$$

Define $\tilde{X}_t = X_t^T v$, $\tilde{\lambda}_t = \lambda_t^T v$ and $\tilde{\mu} = \mu^T v$, where they all belong to $\mathbb{R}^K$. Transposing the above equation we have the topics intensity

$$\tilde{\lambda}_t = \tilde{\mu} + \nu B^T(\phi * d\tilde{X})_t. \tag{3}$$

The intensity $\tilde{\lambda}_t$ of the stochastic process $\tilde{X}_t$ has its k^{th} coordinate given by

$$\tilde{\lambda}_t^k = \sum_{i=1}^{N} \lambda_t^{i,k} v_i, \tag{4}$$

which means that it represents a topic as a weighted sum by users, where the weights are given by each user impact on the social network.

By reference to the previous Obama example: since Obama has assumedly a large v coefficient (he has a large impact on the network), a topic broadcasted by him should be more inclined to be trendy, and thus have a potentially large increase in $\tilde{X}_t$; on the other hand, if a topic is broadcasted by some unknown person, with a small coefficient v, it will almost not affect the topic intensity $\tilde{\lambda}_t$.

[3]This left-eigenvector v^T has all its entries nonnegative, together with the eigenvalue $\nu \geq 0$, by the Perron-Frobenius theorem for matrices with nonnegative entries, without the need of further assumptions. However, we assume without loss of generality that $\nu > 0$, which can be easily avoided during the estimation.

Since $\tilde{X}_t$ is a linear combination of point processes, the increase at time t in $\tilde{X}_t$ can be measured by its intensity $\tilde{\lambda}_t$. *Consequently, we adopt $\tilde{\lambda}^k$ as surrogate for topic k trendiness at time t.*

C. Searching the topic peaks by rescaling

Our algorithm is concerned with the detection of trendy topics at the final diffusion time τ, taking into consideration all the diffusion history in $[0, \tau]$. This means that our goal is to find topics that will possibly have more broadcasts after time τ than they should have, if one looks at their broadcast history in $[0, \tau]$. With that in mind, we say that topic k has a *peak* at time t if its topic intensity $\tilde{\lambda}_t^k$ achieves its maximum expected intensity at time t.

Since the influences $\nu(\phi * d\tilde{X})_t B$ are always nonnegative in Eqn. (3), we can only find peaks when $\tilde{\lambda}^k$ is greater than or equal to its intrinsic mean $\tilde{\mu}^k$. Moreover, one can notice that our definition does not take directly into consideration comparisons between topics, i.e., our definitions of trendiness and of peaks are *relative*, although there exist interactions between topics through the topic-topic influence matrix B.

We continue to the formal derivation of the rescaling, which is performed under the following technical assumption[4]:

Assumption 1. *The topic interaction matrix B can be diagonalized into $B = PDP^{-1}$ (where P is the matrix with the eigenvectors of B and D is a diagonal matrix with the eigenvalues of B) and B has only one maximal eigenvalue.*

Moreover, we assume without loss of generality that $D_{i,i} \geq D_{i+1,i+1}$ and that the largest eigenvalue is $D_{1,1} > 0$ (again, by the Perron-Frobenius theorem, since B has nonnegative entries).

Let us use, for simplicity, exponential kernels, i.e., $\phi(t) = e^{-\omega t}\mathbb{I}_{\{t>0\}}$, where $\omega > 0$ is a parameter that reflects the heaviness of the temporal tail. This means that a larger ω implies a lighter tail, and a smaller temporal interaction between broadcasts.

This choice of kernel function implies that our rescaling uses only one degree of freedom - the timescale parameter ω. It is then quite understandable that with just one degree of freedom we can only have one nontrivial limit behavior for our rescaled topic intensities $\frac{\tilde{\lambda}_{\tau t}^k}{\tau}$. This behavior is thus dictated by the leading eigenvector of B when rescaling. This argument further supports assumption 1.

1) Rescaling the topic intensities: Using the decomposition $B = PDP^{-1}$, where D is a diagonal matrix with the eigenvalues of B, we have that Eqn. (3) can be written as

$$\tilde{\lambda}_t = \tilde{\mu} + \nu(P^{-1})^T D^T P^T(\phi * d\tilde{X})_t,$$

[4]The assumption that B can be diagonalized is in fact a simplifying one. One could use the Jordan blocks of B, on the condition that there exists only one maximal eigenvalue. This assumption is verified if, for example, the graph associated with B is strongly connected; which means that every topic influences the other topics, even if it is in an undirected fashion (by influencing topics that will, in their turn, influence other topics, and so on). One can also develop a theory in the case of multiple maximal eigenvalues for B, but it would be much more complicated as the associated stochastic control problem (as in [21]) has not yet been solved analytically, hence numerical methods should be employed.

which when multiplied by P^T by the left becomes

$$\begin{aligned} P^T\tilde{\lambda}_t &= P^T\tilde{\mu} + \nu D^T P^T(\phi * d\tilde{X})_t \\ &= P^T\tilde{\mu} + \nu D(\phi * P^T d\tilde{X})_t \\ &= P^T\tilde{\mu} + \nu D(\phi * d(P^T\tilde{X}))_t. \end{aligned}$$

Defining $\chi_t = P^T\tilde{X}_t$, $\varphi_t = P^T\tilde{\lambda}_t$ and $\vartheta = P^T\tilde{\mu}$, we have that χ_t is a K-dimensional stochastic process with intensity

$$\varphi_t = \vartheta + \nu D(\phi * d\chi)_t.$$

Under assumption 1, we have

$$\varphi_t^k = \vartheta^k + \nu D_{k,k}(\phi * d\chi^k)_t, \tag{5}$$

where φ_t^k are uncoupled one-dimensional stochastic processes.

Now, following [20], we rescale φ by "pushing" the timescale parameter ω to the unstable regime of $\tilde{X}$, so as to obtain a nontrivial behavior (peak) for the intensity $\tilde{\lambda}$, if any. In light of lemma 1 and assuming an exponential kernel $\phi(t) = e^{-\omega t}.\mathbb{I}_{\{t>0\}}$, we have that the timescale parameter ω satisfies for some $\lambda > 0$

$$\tau(1 - \frac{\nu D_{1,1}}{\omega}) \sim \lambda$$

when $\tau \to \infty$, which implies (we assume without loss of generality that $\tau > \lambda$)

$$\omega \sim \frac{\tau\nu D_{1,1}}{(\tau - \lambda)}. \tag{6}$$

The rescaling stems from the next theorem (the one-dimensional case is the theorem 2.2 in [20]) and is proven in appendix B.

Theorem 1. *Let assumption 1 be true, the temporal kernel be defined as $\phi(t) = e^{-\omega t}.\mathbb{I}_{\{t>0\}}$, let $\rho = ((P^{-1})_{1,1}, \cdots, (P^{-1})_{1,K})$ be the leading left-eigenvector of B, $\tilde{v}$ be the leading right-eigenvector of J, and define $\pi = (\sum_k (P_{k,1})^2\rho_k)(\sum_i v_i^2\tilde{v}_i)$.*

If $\omega \sim \frac{\tau}{\nu D_{1,1}(\tau-\lambda)}$ when $\tau \to \infty$, then the rescaled process $\frac{1}{\tau}\varphi^1_{\tau t}$ converges in law, for the Skorohod[5] topology in $[0,1]$, to a CIR process C^1 satisfying the following stochastic differential equation (SDE)

$$\begin{cases} dC_t^1 = \lambda\nu D_{1,1}(\frac{\vartheta^1}{\lambda} - C_t^1)dt + \nu D_{1,1}\sqrt{\pi}\sqrt{C_t^1}dW_t \\ C_0^1 = 0, \end{cases} \tag{7}$$

where W_t is a standard Brownian motion.

Moreover for $k > 1$, the rescaled processes $\frac{1}{\tau}\varphi^k_{\tau t}$ converge in law to 0 for the Skorohod topology in $[0,1]$.

As a result, we are only interested in the CIR process C^1, since it is the only one that possesses a nontrivial behavior. One can clearly see that, since a CIR process is a mean-reverting one, C^1 mean-reverts to the stationary expectation $\overline{\mu} = \frac{\vartheta^1}{\lambda}$. As already discussed in subsection III-C, if one wants to capture some trend behavior one must see this process above its stationary expectation $\overline{\mu}$, i.e., one must study the process $C_t = C_t^1 - \overline{\mu}$.

By Eqn. (7), one easily has that $C_t = C_t^1 - \overline{\mu}$ satisfies the following SDE:

$$dC_t = -\lambda\nu D_{1,1}C_t dt + \nu D_{1,1}\sqrt{\pi}\sqrt{C_t + \overline{\mu}}dW_t. \tag{8}$$

Remark: A way of pushing the Hawkes process to the instability regime, when estimating the matrices μ, J and B, is to put the timescale parameter ω near the stability boundary given by Eqn. (2).

2) The trend index: After rescaling the $\varphi_t = P^T\tilde{\lambda}_t$, we effectively search for the peaks in $\tilde{\lambda}$ using the framework developed by Espinosa and Touzi [21] dedicated to search for the maximum of scalar mean-reverting Brownian diffusions.

For that goal, we define trend indices $\mathcal{I}_t^k$ as the measure, at each time instant $t \in [0,\tau]$, of how far is the intensity $\tilde{\lambda}^k$ from its peak. To do so, we use the fact that $\tilde{\lambda}_t = (P^{-1})^T\varphi_t$ to determine the limit behavior of $\frac{\tilde{\lambda}^k_{\tau t}}{\tau}$, namely $\tilde{\lambda}_t^{k,\infty}$, as

$$\tilde{\lambda}_t^{k,\infty} = \sum_j P^{-1}_{j,k}C_t^j = P^{-1}_{1,k}C_t^1 = P^{-1}_{1,k}(C_t + \overline{\mu}),$$

where P is the eigenvector matrix of B in assumption 1 and C_t^k are the rescaled CIR processes in theorem 1.

Hence, in order to find our intensity peaks, we consider for each topic k the following optimal stopping problem

$$V_k = \inf_{\theta \in \mathcal{T}_0} \mathbb{E}[\frac{(P^{-1}_{1,k})^2}{2}(C^*_{T_0} - C_\theta)^2], \tag{9}$$

where $C_t^* = \sup_{s \le t} C_s$ is the running maximum of C_t, $T_y = \inf\{t > 0 \mid C_t = y\}$ is the first hitting time of barrier $y \ge 0$ and $\mathcal{T}_0$ is the set of all stopping times θ (with respect to C) such that $\theta \le T_0$ almost surely, i.e., all stopping times until the process C reaches 0.

By the theory developed in [21], one has optimal barriers γ^k relative to each problem V_k. *A barrier represents the peaks of the intensities, i.e., if the CIR process C touches the optimal barrier γ^k, it means that we have found a peak for topic k.*

The authors show that the free barriers γ^k have two monotone parts; first a decreasing part $\gamma^k_\downarrow(x)$ and then an increasing part $\gamma^k_\uparrow(x)$, which are found by solving the ordinary differential equations (ODE) (5.1) and (5.15) in [21], respectively[6].

We are now able to define for each time $t \le T_0$, the temporal trend indices $\mathcal{I}_t^k$ as

$$\mathcal{I}_t^k = \begin{cases} \psi^+(\tau - t, C_t - \gamma^k(C_t)) & \text{if} \quad t < \tau \quad \text{and} \quad C_t \ge 0, \\ \psi^-(\tau - t, C_t - \gamma^k(C_0)) & \text{if} \quad t < \tau \quad \text{and} \quad C_t < 0, \\ \Psi^+(C_\tau - \gamma^k(C_\tau)) & \text{if} \quad t = \tau \quad \text{and} \quad C_t \ge 0, \\ \Psi^-(C_\tau - \gamma^k(C_0)) & \text{if} \quad t = \tau \quad \text{and} \quad C_t < 0, \end{cases}$$

[5] The Skorohod topology in a given space is the natural topology to study càdlàg processes, i.e., stochastic processes that are right-continuous with finite left limits. This topology has the goal to define convergence on cumulative distribution functions and stochastic processes with jumps. See [31] for a formal definition.

[6] For the CIR case we have by Eqn. (8) that the functions α, S and S' defined in [21] are

- $\alpha(x) = \frac{2\lambda x}{\nu D_{1,1}\pi(x+\overline{\mu})}$, $S'(x) = e^{\frac{2\lambda x}{\nu D_{1,1}\pi}}(\frac{x}{\overline{\mu}} + 1)^{-\frac{2\lambda\overline{\mu}}{\nu D_{1,1}\pi}}$ and
- S is a linear combination of a suitable transformation of the confluent hypergeometric functions of first and second kind, M and U, respectively (see [32]), since it must satisfy $S(0) = 0$ and $S'(0) = 1$ (see [33]).

where $\psi^{+/-}$ are decreasing in time (the first variable), increasing in space (the second variable) functions and $\Psi^{+/-}$ are increasing in space functions. We impose $\psi^{+/-}$ as decreasing functions of time because our trend detection algorithm is to determine the trendy topics at time τ, the end of the estimation time period. Thus the further we are in the past (measured by $\tau - t$), the less influence it must have in our decision, and consequently in our trend index. By the same token, $\psi^{+/-}$ and $\Psi^{+/-}$ must be increasing functions in space because we want to distinguish topics that have higher intensities, and penalize those that have a lower intensity, thus if the intensity is bigger than the optimal barrier, we must give it a bigger index. If, on the other hand, the intensity is smaller than the optimal barrier, even negative in some cases, we must take into account the degree of this separation. One has the liberty to choose the functions ψ and Ψ according to some calibration dataset, which makes the model more versatile and data-driven.

Please note that in the definition of $\mathcal{I}_t^k$, the following factors have been taken into consideration:

- even if the CIR intensity C_t did not reach its expected maximum given by $\gamma^k(C_t)$, we must account for the fact that it may have been close enough,
- reaching the expected maximum is good, but surpassing it is even better. So we must not only define a high trend index if C_t reaches the expected maximum given by $\gamma^k(C_t)$, but we must define a *higher* trend index if C_t surpasses these barriers, and
- it is important to *penalize* all the times $t \in [0, \tau]$ that the intensity C_t becomes negative, i.e., the intensities $\tilde{\lambda}_t^k$ become smaller than their stationary expectation.

The trend indices $\mathcal{I}^k$ are thus defined as

$$\mathcal{I}^k = \int_0^\tau \mathcal{I}_t^k dt.$$

Remark: One could be also interested in not only tracking the relative trendiness of each topic with respect to their maxima, but also the *absolute* trendiness of topics with respect to each other. In this case, one may define the trend indices $\tilde{\mathcal{I}}_t^k$ as

$$\tilde{\mathcal{I}}_t^k = \mathcal{I}_t^k + a(\tau - t)\tilde{\lambda}_t^{k,\infty} = \mathcal{I}_t^k + a(\tau - t)P_{1,k}^{-1}(C_t + \overline{\mu}),$$

where $a(\tau - t) \geq 0$ are nonincreasing functions of time (again, in order to give a bigger influence to the present compared to the past). The absolute trendiness of topics can be explained as follows: Lady Gaga may be not trendy according to our definition, if for example people do not tweet *as much as expected* about her at the moment, but she will probably still be trendier than a rising-but-still-obscure Punk-Rock band. In this case, the relative trend index $\mathcal{I}^k$ of Lady Gaga is not that big as compared to the relative trend index of the Punk-Rock band. However, the absolute trend index $\tilde{\mathcal{I}}^k$ of Lady Gaga will surely be bigger than the absolute trend index of the Punk-Rock band, if the function $a(\tau - t)$ is large enough. The function $a(\tau - t)$ controls which behavior one wants to detect, the relative or the absolute trendiness.

Remark: This algorithm is fast, despite the use of numerical discretization schemes for the ODEs. By using the eigenvector centrality of the underlying social network as tool to create our trend indices, we not only use the topological properties of the social network in question but we reduce considerably the dimension of the problem: we only have a one-dimensional CIR process to study. Moreover, the complexity of the algorithm breaks down to three parts: 1) the resolution of the K optimal barrier ODEs, which is of order $\mathcal{O}(\frac{K}{\delta})$ where δ is the time-discretization step, 2) the calculation of the left and right leading eigenvectors of J and B, which can be achieved fairly fast with iterative methods such as the power method, and 3) the matrix product in the calculation of $\overline{\mu}$, which has complexity $\mathcal{O}(NK)$.

Algorithm 1 Trend detection algorithm

Input: Hawkes process $X_t,\ t \in [0, \tau]$, matrices J, B and μ
1**:** Compute the leading left-eigenvector v^T and eigenvalue ν of J, and the topic intensities $\tilde{\lambda}_t$ following Eqn. (3)
2**:** Compute the leading right-eigenvector $(P_{11}, \cdots, P_{K1})$, left-eigenvector $(P_{11}^{-1}, \cdots, P_{1K}^{-1})$ and eigenvalue $D_{1,1}$ of B, and the leading right-eigenvector $\tilde{v}$ of J
3**:** "Push" $\tilde{\lambda}_t$ to the instability regime following Eqn. (6) and calculate the CIR intensity C_t following Eqn. (8)
4**:** Discretize $[0, \tau]$ into T bins of size $\delta \ll 1$
for $k = 1$ **to** K **do**
    5**:** Get the optimal barrier γ^k in $\{0, \delta, 2\delta, \cdots, (T-1)\delta\}$, following [21]
    for $t = 1$ **to** T **do**
        5**:** Calculate the trend index $\mathcal{I}_{(t-1)\delta}^k$ using the optimal barrier γ^k of the optimal stopping problem (9)
    end for
    6**:** Calculate the topic trend index $\mathcal{I}^k = \int_0^\tau \mathcal{I}_t^k dt = \delta \sum_{t=1}^T \mathcal{I}_{(t-1)\delta}^k$
end for
Output: Trend indices $\mathcal{I}^k$

IV. Numerical Examples

We provide in this section two numerical examples of our trend detection algorithm. The first example (which we term SIM) is a synthetic near unstable Hawkes processes in a social network using Ogata's thinning algorithm[7] [23] in a time horizon $\tau = 50$. We used 10 topics for the simulation, the last 5 topics not possessing *any topic influence*, i.e., $B_{c,k} = 0$ for all c and $k \in \{6, 7, 8, 9, 10\}$, corresponding to figures 1, 2 and 3. The second example is a MemeTracker dataset (MT), with different topics and world news for the $5,000$ most active sites from 4 million sites from March 2011 to February 2012[8]. We used the 5 most and 5 least broadcasted memes and a maximum likelihood estimation procedure for the parameters, leading to figures 4, 5 and 6.

For both datasets, Figures 1 and 4 plot the scaled topic intensities $\frac{\tilde{\lambda}_{\tau t}}{\tau}$, Figures 2 and 5 plot the cumulative number of broadcasts of each topic, i.e., $\overline{X}_t^k = \sum_i X_t^{i,k}$, and Figures 3

[7] The thinning algorithm simulates a standard Poisson process P_t with intensity $M > \sum_{i,k} \lambda_t^{i,k}$ for all $t \in [0, \tau]$ and selects from each jump of P_t the Hawkes jumps of $X_t^{i,k}$ with probability $\frac{\lambda_t^{i,k}}{M}$, or no jump at all with probability $\frac{M - \sum_{i,k} \lambda_t^{i,k}}{M}$.

[8] Data available at http://snap.stanford.edu/netinf.

and 6 plot the optimal free barriers $\gamma^k(C_t)$ at time t, for each topic k, against the current maximum of C_t.

We observe that in figures 3 and 6 the CIR processes touch the barriers, which means that the topic intensities indeed reached their peaks.

We compute in table I the trend indices $\tilde{\mathcal{I}}^k$ for both datasets. We used for the trend indices calculation the following functions $\psi^{+/-}(t,x) = \frac{e^{2x}}{t+1}$, $\Psi^{+/-}(x) = 2x$ and $a(t) = \frac{1}{t+1}$, as explained in subsubsection III-C2.

In reference to table 1, one can see that the trend index for topic 5 is the highest in both datasets, however for different reasons: in the synthetic dataset, this is due to the fact that topic 5 has the largest topic intensity. But for the MemeTracker dataset, topic 5 shows higher trendiness than the other topics because it has a peak in intensity later in time, even though it does not have the highest topic intensity. *This phenomenon illustrates the difference between our algorithm and one that looks solely to the largest topic intensities.*

In addition, one can see in figures 2 and 5 that topic 5 does not have the largest number of jumps in both datasets, even though it is the trendiest by table I. This result shows that the number of messages is not the only factor in our algorithm, it takes equally into account the social ties and influences between users and topics in the social network.

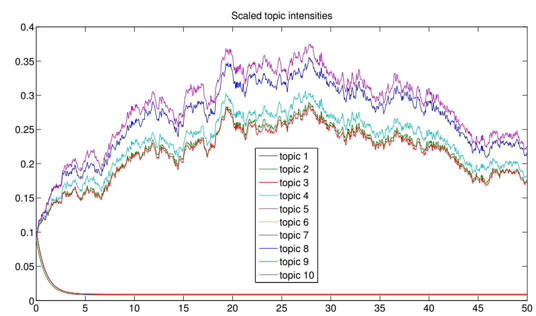

Fig. 1. Topic intensities $\tilde{\lambda}_t = \sum_i \lambda_t^{i,k} v_i$ for the synthetic dataset.

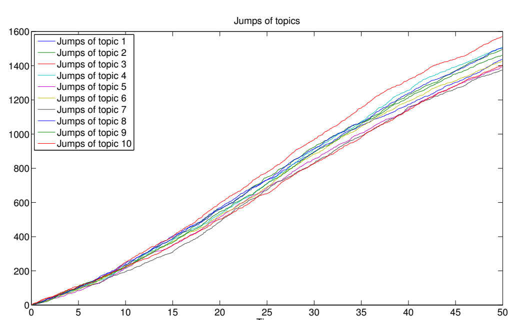

Fig. 2. Cumulative sum of jumps of topics $\overline{X}_t = \sum_i X_t^{i,k}$ for the synthetic dataset.

V. Conclusion

We have developed in this paper a trend detection algorithm, designed to find trendy topics being disseminated in a social network. We have assumed that broadcasts of

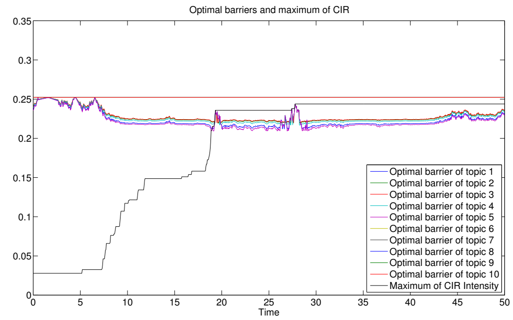

Fig. 3. Maximum of CIR process, C_t^*, plotted against the optimal barriers $\gamma^k(C_t)$ for each topic k for the synthetic dataset

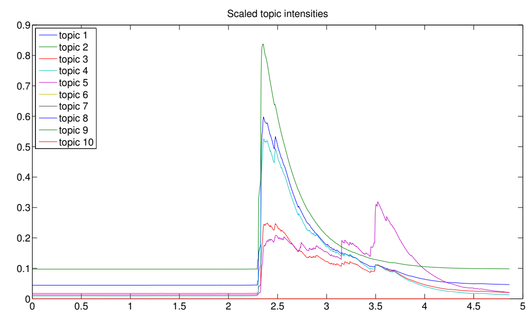

Fig. 4. Topic intensities $\tilde{\lambda}_t = \sum_i \lambda_t^{i,k} v_i$ for the meme tracker dataset.

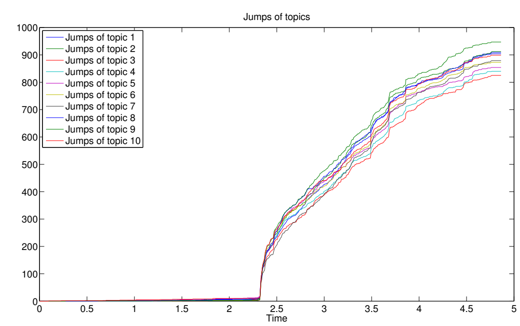

Fig. 5. Cumulative sum of jumps of topics $\overline{X}_t = \sum_i X_t^{i,k}$ for meme tracker dataset.

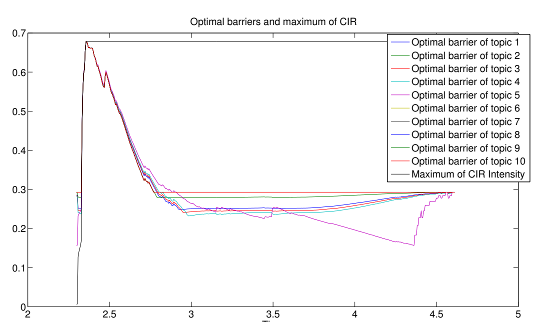

Fig. 6. Maximum of CIR process, C_t^*, plotted against the optimal barriers $\gamma^k(C_t)$ for each topic k in the meme tracker dataset.

messages in the social network can be modeled by a self-

TABLE I. TREND INDICES FOR EACH DATASET.

DATASET	$\tilde{\mathcal{I}}^1$	$\tilde{\mathcal{I}}^2$	$\tilde{\mathcal{I}}^3$	$\tilde{\mathcal{I}}^4$	$\tilde{\mathcal{I}}^5$	$\tilde{\mathcal{I}}^6$	$\tilde{\mathcal{I}}^7$	$\tilde{\mathcal{I}}^8$	$\tilde{\mathcal{I}}^9$	$\tilde{\mathcal{I}}^{10}$
SIM	0.0481	0.0414	0.0407	0.0433	0.0507	0.0148	0.0148	0.0148	0.0148	0.0148
MT	0.1175	0.1090	0.1199	0.1229	0.1404	0.1035	0.1035	0.1035	0.1035	0.1035

exciting point process, namely a Hawkes process, which takes into consideration the real broadcasting times of messages and the interaction between users and topics.

We defined our idea of trendiness and derived trend indices for each topic being disseminated. These indices take into consideration the time between the actual trend detection and the message broadcasts, the distance between the intensity of broadcasting and the maximum expected intensity of broadcasting, and the social network topology. This result is, to the best of our knowledge, the first definition of relative trendiness, i.e., a topic may not be very trendy in absolute number of broadcasts when compared to other topics, but has still rapid and significant number of broadcasts as compared to its expected behavior. Still, one can easily create an absolute trend index for each topic in our trend detection algorithm, where all one needs to do is use the broadcasting intensities of each topic as surrogates for their trendiness. It is worthy mentioning that these broadcast intensities also take into consideration the social network topology, or more precisely, the outgoing eigenvector centrality of each user, i.e., their respective influences on the social network.

The proposed trend detection algorithm is simple and uses stochastic control techniques in order to derive a free barrier for the maximum expected broadcast intensity of each topic. This method is fast and aggregates all the information of the point process into a simple one-dimensional diffusion, thus reducing its complexity and the quantity of data necessary to the detection - indispensable features if one is concerned with the detection of trends in real-life social networks.

APPENDIX A
ASSUMPTIONS AND MAIN THEOREM

We now proceed to the proof of theorem 1, following the ideas in [20]. According to section II, we have a multivariate linear Hawkes process $X_t^{i,k}$ with intensity of the form

$$\lambda_t^{i,k} = \mu^{i,k} + \sum_c \sum_j B_{c,k} J_{i,j} \int_0^{t-} \phi(t-s) dX_s^{j,c}, \qquad (10)$$

which in matrix form can be seen as

$$\lambda_t = \mu + J(\phi * dX)_t B,$$

where μ is the intrinsic rate of dissemination, J is the user-user interaction matrix, B is the topic-topic interaction matrix and $(\phi * dX)_t$ is the $N \times K$ convolution matrix defined as $(\phi * dX)_t^{i,k} = \int_0^t \phi(t-s) dX_s^{i,k}$.

In order to prove our main rescaling convergence result, we make the following assumptions:

Assumption 2. *The temporal kernel $\phi(t)$ is an exponential function with timescale parameter ω_τ*

$$\phi(t) = e^{-\omega_\tau t}.\mathbb{I}_{\{t>0\}}.$$

Remark: Assumption 2 is in fact a simplifying one, and one may use any temporal kernel satisfying the hypothesis in [20].

Assumption 3. *The interaction matrices J and B can be diagonalized into $J = v^{-1}\nu v$ and $B = \rho D \rho^{-1}$ and $B^T \otimes J$ has only one maximal eigenvalue. Thus, in light of the decomposition for J and B, we have that J has left-eigenvectors the rows of v, denoted by v_i^T, with associated eigenvalues ν_i; and B has right-eigenvectors the columns of ρ, denoted by ρ_k, with associated eigenvalues $D_{k,k}$, i.e., v^T is the $N \times N$ matrix and ρ is the $K \times K$ matrix*

$$v^T = \begin{pmatrix} | & \cdots & | \\ v_1^T & \cdots & v_N^T \\ | & \cdots & | \end{pmatrix} \quad \textit{and} \quad \rho = \begin{pmatrix} | & \cdots & | \\ \rho_1 & \cdots & \rho_K \\ | & \cdots & | \end{pmatrix}.$$

Since the eigenvalues of $B^T \otimes J$ are of the form $v_i D_{k,k}$, (i,k), we assume without loss of generality that $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_N$ and $D_{1,1} > D_{2,2} \geq D_{3,3} \geq \cdots \geq D_{K,K}$, and that the largest eigenvalues of J and B satisfy $\nu_1 > 0$ and $D_{1,1} > 0$.

Moreover, we also have that v_1^T and ρ_1 have nonnegative entries by the Perron-Frobenius theorem, since J and B have nonnegative entries (this result remains true for the leading right-eigenvector of J and the leading left-eigenvector of B as well).

Remark: The assumption that J and B can be diagonalized is in fact a simplifying one. One could use the Jordan blocks of J and B, on the condition that there exists only one maximal eigenvalue for $B^T \otimes J$. This assumption is verified if, for example, the graph associated with B is strongly connected; which means that every topic influences the other topics, even if it is in an undirected fashion (by influencing topics that will, in their turn, influence other topics, and so on). One can also develop a theory in the case of multiple maximal eigenvalues, but it will be much more complicated and the associated stochastic control problem has not yet be solved analytically, hence numerical methods must be employed.

Assumption 4. *We have that the timescale parameter ω_τ satisfies, for some $\lambda > 0$,*

$$\tau\left(1 - \frac{\nu_1 D_{1,1}}{\omega_\tau}\right) \to \lambda$$

when $\tau \to \infty$, which implies

$$\omega_\tau \searrow \nu_1 D_{1,1}.$$

Appendix B is thus responsible for the proof of the following theorem:

Theorem 2. *Let X be the multivariate Hawkes process in $[0, \tau]$ with intensity given by Eqn. (10), and let $\varphi_t^{i,k} = v_i^T \frac{\lambda_{\tau t}}{\tau} \rho_k$, where v_i^T and ρ_k are defined in assumption 3.*

Under assumptions 2, 3 and 4 we have that

- *If $(i,k) \neq (1,1)$ then $\varphi_t^{i,k}$ converges in law to 0 for the Skorokhod topology in $[0,1]$ when $\tau \to \infty$.*
- *Let v_1^T and $\tilde{v}_1$ be the leading left and right eigenvectors of J associated with the eigenvalue $\nu_1 > 0$, let ρ_1 and $\tilde{\rho}_1^T$ be the leading right and left eigenvectors of B associated with the eigenvalue $D_{1,1} > 0$, and define $\pi = (\sum_i v_{1,i}^2 \tilde{v}_{i,1})(\sum_k \rho_{k,1}^2 \tilde{\rho}_{1,k})$.*
  Thus, $\varphi_t^{1,1}$ converges in law to the CIR process C_t for the Skorokhod topology in $[0,1]$ when $\tau \to \infty$, where $C_t,\ t \in [0,1]$ satisfies the following stochastic differential equation
  $$\begin{cases} dC_t &= \lambda \nu_1 D_{1,1}(\frac{\mu}{\lambda} - C_t)dt + \nu_1 D_{1,1}\sqrt{\pi}\sqrt{C_t}dW_t, \\ C_0 &= 0, \end{cases}$$
  where W_t is a standard Brownian motion.

Appendix B
Proof of theorem 2

A. Sketch of proof

We provide here a sketch of the proof:

1) We start by writing the equations satisfied by the rescaled intensities $\varphi_t^{i,k} = v_i^T \frac{\lambda_{\tau t}}{\tau} \rho_k$ and study their first-order properties.
2) Secondly, we define the new martingales $B_t^{i,k}$ and show that they converge to a standard Brownian motion.
3) Thirdly, we rewrite $\varphi_t^{1,1}$ in a more suitable form, with remainder terms U_t and V_t, and we show that they converge to 0.
4) Finally, we apply the convergence theorem 5.4 of [34] for limits of stochastic integrals with semimartingales.

B. Rescaling the Hawkes intensity

Let us begin by defining the one-dimensional stochastic processes
$$\tilde{\lambda}_t^{i,k} = v_i^T \lambda_t \rho_k,$$
which satisfy the one-dimensional equations
$$\begin{aligned}\tilde{\lambda}_t^{i,k} &= v_i^T \mu \rho_k + v_i^T J(\phi * dX)_t B \rho_k \\ &= \tilde{\mu}^{i,k} + \nu_i D_{k,k}(\phi * \tilde{\lambda}^{i,k})_t + \nu_i D_{k,k}(\phi * v_i^T dM \rho_k)_t,\end{aligned} \tag{11}$$
with $M_t = X_t - \int_0^t \lambda_s ds$ the compensated martingale associated with the Hawkes process X and $\tilde{\mu}^{i,k} = v_i^T \mu \rho_k$. Using lemma 2.1 of [20], we have that
$$\tilde{\lambda}_t^{i,k} = \tilde{\mu}^{i,k} + \tilde{\mu}^{i,k}\int_0^t \Psi_{i,k}(t-s)ds + \int_0^t \Psi_{i,k}(t-s) v_i^T dM_s \rho_k, \tag{12}$$
where
$$\Psi_{i,k}(t) = \sum_{n \geq 1} (\nu_i D_{k,k}\phi(t))^{*n},$$
with the n^{th} convolution operator defined as
$$\begin{aligned}&(\nu_i D_{k,k}\phi(t))^{*1} = \nu_i D_{k,k}\phi(t), \quad \text{and} \\ &(\nu_i D_{k,k}\phi(t))^{*n} = ((\nu_i D_{k,k}\phi)^{*(n-1)} * \nu_i D_{k,k}\phi)_t.\end{aligned}$$

We have the following lemma for the convolutions $\Psi_{i,k}$:

Lemma 2. *Let $\Psi_{i,k}(t) = \sum_{n\geq 1}(\nu_i D_{k,k}\phi(t))^{*n}$, then under assumption 2 we have that*
$$\Psi_{i,k}(t) = \nu_i D_{k,k} e^{-\omega_\tau(1 - \frac{\nu_i D_{k,k}}{\omega_\tau})t}.$$

Moreover, under assumptions 3 and 4 we have that
$$\Psi_{1,1}(\tau t) \to \nu_1 D_{1,1} e^{-\nu_1 D_{1,1}\lambda t}$$
uniformly in $[0,1]$ when $\tau \to \infty$, and that there exists a constant $L > 0$ such that for $(i,k) \neq (1,1)$ we have
$$\int_0^t \Psi_{i,k}(\tau(t-s))ds \leq \frac{L}{\tau}. \tag{13}$$

Proof: Under assumption 2, we have that
$$\begin{aligned}(\nu_i D_{k,k}\phi(t))^{*2} &= (\nu_i D_{k,k})^2 \int_0^t e^{-\omega_\tau(t-s)} e^{-\omega_\tau s} ds \\ &= (\nu_i D_{k,k})^2 t e^{-\omega_\tau t} \\ &\Rightarrow (\nu_i D_{k,k}\phi(t))^{*n} = (\nu_i D_{k,k})^n \frac{t^{n-1}}{(n-1)!} e^{-\omega_\tau t},\end{aligned}$$
hence
$$\Psi_{i,k}(t) = e^{-\omega_\tau t}\sum_{n\geq 1} \nu_i^n D_{k,k}^n \frac{t^{n-1}}{(n-1)!} = \nu_i D_{k,k} e^{-(1 - \frac{\nu_i D_{k,k}}{\omega_\tau})\omega_\tau t}.$$

Now, under assumptions 3 and 4, we have that $\tau(1 - \frac{\nu_1 D_{1,1}}{\omega_\tau}) \to \lambda$ and $\omega_\tau \to \nu_1 D_{1,1}$, which implies that there exists a constant $\underline{\lambda} > 0$ such that for every $(i,k) \neq (1,1)$
$$\omega_\tau(1 - \frac{\nu_i D_{k,k}}{\omega_\tau}) \geq \underline{\lambda} > 0 \quad \text{and} \quad \tau(1 - \frac{\nu_i D_{k,k}}{\omega_\tau}) \to \infty.$$

Firstly, for $t \in [0,1]$, using the Lipschitz continuity of e^{-t} we have that
$$\begin{aligned}&|\Psi_{1,1}(\tau t) - \nu_1 D_{1,1} e^{-\nu_1 D_{1,1}\lambda t}| \\ &\leq (\nu_1 D_{1,1})^2 \lambda (\sup_{s\in[0,1]} e^{-\nu_1 D_{1,1}\lambda s}) t |\omega_\tau \tau(1 - \frac{\nu_1 D_{1,1}}{\omega_\tau}) - \nu_1 D_{1,1}\lambda| \\ &\leq (\nu_1 D_{1,1})^2 \lambda |\omega_\tau \tau(1 - \frac{\nu_1 D_{1,1}}{\omega_\tau}) - \nu_1 D_{1,1}\lambda| \to 0\end{aligned}$$
when $\tau \to \infty$, which implies that $\Psi_{1,1}(\tau t) \to \nu_1 D_{1,1} e^{-\nu_1 D_{1,1}\lambda t}$ uniformly in $[0,1]$.

At last, for $(i,k) \neq (1,1)$, we have that
$$\int_0^t \Psi_{i,k}(\tau(t-s))ds = \nu_i D_{k,k} \frac{1 - e^{-t\tau\omega_\tau(1 - \frac{\nu_i D_{k,k}}{\omega_\tau})}}{\tau\omega_\tau(1 - \frac{\nu_i D_{k,k}}{\omega_\tau})} \leq \frac{L}{\tau}$$
where $L > 0$ is a large enough positive constant. ■

Let us now define the one-dimensional rescaled stochastic processes, for $t \in [0,1]$,
$$\varphi_t^{i,k} = \frac{v_i^T \lambda_{\tau t}\rho_k}{\tau},$$

which clearly satisfies

$$\varphi_t = v\frac{\lambda_{\tau t}}{\tau}\rho. \tag{14}$$

We have the following lemma concerning the first order properties of φ_t:

Lemma 3. *Let us define the $1\times N$ row vector $v_i^{\odot 2}$ such that $(v_i^{\odot 2})_j = v_{i,j}^2$ and the $K\times 1$ vector $\rho_k^{\odot 2}$ such that $(\rho_k^{\odot 2})_c = \rho_{c,k}^2$. We have*

1) *$\varphi_t^{i,k}$ satisfies the following equation*

$$\varphi_t^{i,k} = \tilde{\mu}^{i,k}\Big(\frac{1}{\tau} + \int_0^t \Psi_{i,k}(\tau(t-s))ds\Big) + \int_0^t \Psi_{i,k}(\tau(t-s))\sqrt{(v_i^{\odot 2})^T\frac{\lambda_{\tau s}}{\tau}\rho_k^{\odot 2}}dB_s^{i,k}, \tag{15}$$

where

$$B_t^{i,k} = \sqrt{\tau}\int_0^t \frac{v_i^T dM_{\tau s}\rho_k}{\sqrt{(v_i^{\odot 2})^T\lambda_{\tau s}\rho_k^{\odot 2}}} = \int_0^t \frac{v_i^T dM_{\tau s}\rho_k}{\sqrt{(v_i^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_k^{\odot 2}}}$$

is a L^2 martingale.

2) *If $(i,k)\neq(1,1)$, then*

$$\mathbb{E}[\varphi_t^{i,k}] \le \frac{L}{\tau},$$

where $L>0$ is a large enough positive constant. Moreover, we also have that

$$\mathbb{E}[\varphi_t^{1,1}] \le L.$$

Proof:

1) We have that

$$\begin{aligned}
\varphi_t^{i,k} &= \frac{1}{\tau}\tilde{\mu}^{i,k} + \frac{1}{\tau}\tilde{\mu}^{i,k}\int_0^{\tau t}\Psi_{i,k}(\tau t - s)ds \\
&+ \frac{1}{\tau}\int_0^{\tau t}\Psi_{i,k}(\tau t - s)v_i^T dM_s\rho_k \\
&= \frac{1}{\tau}\tilde{\mu}^{i,k} + \tilde{\mu}^{i,k}\int_0^t \Psi_{i,k}(\tau(t-s))ds \\
&+ \int_0^t \Psi_{i,k}(\tau(t-s))v_i^T dM_{\tau s}\rho_k \\
&= \frac{1}{\tau}\tilde{\mu}^{i,k} + \tilde{\mu}^{i,k}\int_0^t \Psi_{i,k}(\tau(t-s))ds \\
&+ \int_0^t \Psi_{i,k}(\tau(t-s))\sqrt{(v_i^{\odot 2})^T\frac{\lambda_{\tau s}}{\tau}{}_s\rho_k^{\odot 2}}\frac{v_i^T\sqrt{\tau}dM_{\tau s}\rho_k}{\sqrt{(v_i^{\odot 2})^T\lambda_{\tau s}\rho_k^{\odot 2}}} \\
&= \frac{1}{\tau}\tilde{\mu}^{i,k} + \tilde{\mu}^{i,k}\int_0^t \Psi_{i,k}(\tau(t-s))ds \\
&+ \int_0^t \Psi_{i,k}(\tau(t-s))\sqrt{(v_i^{\odot 2})^T\frac{\lambda_{\tau s}}{\tau}{}_s\rho_k^{\odot 2}}dB_s^{i,k}.
\end{aligned}$$

As $\frac{\lambda_{\tau s}}{\tau} = v^{-1}\varphi_s\rho^{-1}$ by Eqn. (14), we have the result.

2) Since $B_t^{i,k}$ is a martingale, we have that

$$\begin{aligned}
\mathbb{E}[\varphi_t^{i,k}] &= \frac{1}{\tau}\tilde{\mu}^{i,k} + \frac{1}{\tau}\tilde{\mu}^{i,k}\int_0^{\tau t}\Psi_{i,k}(\tau t - s)ds \\
&= \frac{1}{\tau}\tilde{\mu}^{i,k} + \tilde{\mu}^{i,k}\int_0^t \Psi_{i,k}(\tau(t-s))ds,
\end{aligned}$$

which together with lemma 2 gives us the result.

■

Remark: We can assume, without loss of generality, that there exists a $c>0$ such that $\min_{(i,k)} v_i^T\mu\rho_k = \min_{(i,k)}\tilde{\mu}^{i,k} \ge c$, since $v_i^T\mu\rho_k = \tilde{\mu}^{i,k} = 0 \Rightarrow \mathbb{E}[\varphi_t^{i,k}] = 0$ by lemma 3, which implies $\varphi_t^{i,k} = 0$ almost surely for all $t\ge 0$ by the fact that $\varphi_t^{i,k}\ge 0$ for all $t\ge 0$.

C. Second order properties

Regarding the second order properties of $B_t^{i,k}$ and $\varphi_t^{i,k}$, we have the following lemma:

Lemma 4. *For each (i,k) let $[B^{i,k}]_t$ be the quadratic variation of the martingale $B_t^{i,k}$. We have that*

1)
$$[B^{i,k}]_t = t + \frac{1}{\tau}\int_0^{\tau t}\frac{(v_i^{\odot 2})^T dM_s\rho_k^{\odot 2}}{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}}. \tag{16}$$

2)
$$\mathbb{E}[(\varphi_t^{i,k})^2] \le L \tag{17}$$

for a constant $L>0$.

3) *Moreover, if $(i,k)\neq(1,1)$, then*

$$\mathbb{E}[(\varphi_t^{i,k})^2] \le \frac{L}{\tau^2}$$

for a constant $L>0$.

Proof:

1) Since the Hawkes process X does not have more than one jump at each time, we have that

$$[M^{i,k}, M^{j,c}]_t = X_t^{i,k}\mathbb{I}_{\{(i,k)=(j,c)\}},$$

which by its turn implies

$$[v_i^T M\rho_k]_t = (v_i^{\odot 2})^T X_t\rho_k^{\odot 2}.$$

We have by lemma 3 that

$$B_t^{i,k} = \sqrt{\tau}\int_0^t \frac{v_i^T dM_{\tau s}\rho_k}{\sqrt{(v_i^{\odot 2})^T\lambda_{\tau s}\rho_k^{\odot 2}}} = \frac{1}{\sqrt{\tau}}\int_0^{\tau t}\frac{v_i^T dM_s\rho_k}{\sqrt{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}}},$$

hence

$$\begin{aligned}
[B^{i,k}]_t &= \frac{1}{\tau}\int_0^{\tau t}\frac{(v_i^{\odot 2})^T dX_s\rho_k^{\odot 2}}{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}} \\
&= \frac{1}{\tau}\int_0^{\tau t}\frac{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}}{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}}ds + \frac{1}{\tau}\int_0^{\tau t}\frac{(v_i^{\odot 2})^T dM_s\rho_k^{\odot 2}}{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}} \\
&= t + \frac{1}{\tau}\int_0^{\tau t}\frac{(v_i^{\odot 2})^T dM_s\rho_k^{\odot 2}}{(v_i^{\odot 2})^T\lambda_s\rho_k^{\odot 2}}.
\end{aligned}$$

2) Using the fact that $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we have by Eqn. (15) and lemma 2 that

$$(\varphi_t^{i,k})^2 \leq 3\Big(\frac{(\tilde{\mu}^{i,k})^2}{\tau^2} + (\tilde{\mu}^{i,k})^2 (\int_0^t \Psi_{i,k}(\tau(t-s))ds)^2 + (\int_0^t \Psi_{i,k}(\tau(t-s))\sqrt{(v_i^{\odot 2})^T v^{-1}\varphi_s \rho^{-1}\rho_k^{\odot 2}} dB_s^{i,k})^2\Big)$$
$$\leq L' + 3(\int_0^t \Psi_{i,k}(\tau(t-s)) \sqrt{(v_i^{\odot 2})^T v^{-1}\varphi_s \rho^{-1}\rho_k^{\odot 2}} dB_s^{i,k})^2.$$

Since $\Psi_{i,k}(\tau(t-s)) = \Psi_{i,k}(\tau t)\Psi_{i,k}(-\tau s)$, we have then

$$(\varphi_t^{i,k})^2 \leq L' + 3(\int_0^t \Psi_{i,k}(\tau(t-s)) \sqrt{(v_i^{\odot 2})^T v^{-1}\varphi_s \rho^{-1}\rho_k^{\odot 2}} dB_s^{i,k})^2 = L' + 3\Psi_{i,k}^2(\tau t)(Z_t^{i,k})^2,$$

with $Z_t^{i,k} = \int_0^t \Psi_{i,k}(-\tau s)\sqrt{(v_i^{\odot 2})^T v^{-1}\varphi_s \rho^{-1}\rho_k^{\odot 2}} dB_s^{i,k}$ a martingale. By lemmas 8 and 3 we have that

$$\mathbb{E}[(\varphi_t^{i,k})^2] \leq L' + 3\Psi_{i,k}^2(\tau t)\int_0^t \Psi_{i,k}^2(-\tau s) (v_i^{\odot 2})^T v^{-1}\mathbb{E}[\varphi_s]\rho^{-1}\rho_k^{\odot 2} ds$$
$$= L' + 3\int_0^t \Psi_{i,k}^2(\tau(t-s)) (v_i^{\odot 2})^T v^{-1}\mathbb{E}[\varphi_s]\rho^{-1}\rho_k^{\odot 2} ds \leq L.$$

3) For $(i,k) \neq (1,1)$, we have from Eqn. (12), lemma 2 and the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ that

$$(\tilde{\lambda}_t^{i,k})^2 \leq L' + 3(\int_0^t \Psi_{i,k}(t-s) v_i^T dM_s \rho_k)^2.$$

One can promptly see by lemma 3 that $\mathbb{E}[\lambda_t^{i,k}] = (v^{-1}\mathbb{E}[\tilde{\lambda}_t]\rho^{-1})_{i,k} \leq L'''$. Hence using lemma 8 and the same calculation of the previous item gives

$$\mathbb{E}[(\tilde{\lambda}_t^{i,k})^2] \leq L' + 3\int_0^t \Psi_{i,k}^2(t-s)(v_i^{\odot 2})^T \mathbb{E}[\lambda_s]\rho_k^{\odot 2} ds$$
$$\leq L''(1 + \int_0^t \Psi_{i,k}^2(t-s)ds) \leq L$$

for a constant $L > 0$. Thus $\mathbb{E}[(\varphi_t^{i,k})^2] = \frac{\mathbb{E}[(\tilde{\lambda}_{\tau t}^{i,k})^2]}{\tau^2} \leq \frac{L}{\tau^2}$, as desired.

■

We derive next the convergence properties of the martingales $B_t^{i,k}$ and the rescaled process $\varphi_t^{i,k}$, $(i,k) \neq (1,1)$.

Lemma 5. *We have that*

1) *For every (i,k), $B_t^{i,k}$ converges in law to a standard Brownian motion for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.*
2) *If $(i,k) \neq (1,1)$, then $\varphi_t^{i,k}$ converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.*

Proof:

1) By Eqn. (16) we have for $t \in [0,1]$

$$\mathbb{E}[([B^{i,k}]_t - t)^2] = \mathbb{E}[(\frac{1}{\tau}\int_0^{\tau t} \frac{(v_i^{\odot 2})^T dM_s \rho_k^{\odot 2}}{(v_i^{\odot 2})^T \lambda_s \rho_k^{\odot 2}})^2]$$
$$\leq \frac{1}{\tau^2}\mathbb{E}[\int_0^{\tau t} \frac{d[(v_i^{\odot 2})^T M \rho_k^{\odot 2}]_s}{\left((v_i^{\odot 2})^T \lambda_s \rho_k^{\odot 2}\right)^2}]$$
$$= \frac{1}{\tau^2}\mathbb{E}[\int_0^{\tau t} \frac{(v_i^{\odot 4})^T \lambda_s \rho_k^{\odot 4} ds}{\left((v_i^{\odot 2})^T \lambda_s \rho_k^{\odot 2}\right)^2}]$$
$$\leq ||v_i^{\odot 2}||.||\rho_k^{\odot 2}||.\frac{1}{\tau^2}\mathbb{E}[\int_0^{\tau t} \frac{ds}{(v_i^{\odot 2})^T \lambda_s \rho_k^{\odot 2}}]$$
$$\leq L\frac{1}{\tau^2}\int_0^{\tau t} dt \leq \frac{L}{\tau}$$

for some $L > 0$ by lemma 9.
Thus, by Markov's inequality we have that for all $\varepsilon > 0$ and for all $t \in [0,1]$

$$\mathbb{P}(|[B^{i,k}]_t - t| \geq \varepsilon) \leq \frac{L}{\tau\varepsilon^2} \to 0 \ \text{ when } \ \tau \to \infty,$$

which shows that, for every $t \in [0,1]$, $[B^{i,k}]_t$ converges in probability towards t when $\tau \to \infty$.
Since $B^{i,k}$ has uniformly bounded jumps because X and λ have uniformly bounded jumps, we have by theorem $VIII.3.11$ of [35] that $B_t^{i,k}$ converges in law to a standard Brownian motion for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.

2) Since $\sup_{t\in[0,1]} \mathbb{E}[\varphi_t^{i,k}] \to 0$ when $\tau \to \infty$ by lemma 3, we have by Eqn. (15) that we only need to prove the convergence of $Z_t^{i,k} = \int_0^t \Psi_{i,k}(\tau(t-s))g(\varphi_s)dB_s^{i,k}$, where $g(\varphi_s) = \sqrt{(v_i^{\odot 2})^T v^{-1}\varphi_s \rho^{-1}\rho_k^{\odot 2}}$ satisfies $|g(\varphi_s)| \leq C(1 + ||\varphi_s||)$ for some $C > 0$.
Since $\Psi_{i,k}(\tau(t-s))$ is an exponential function by lemma 2, we have that assumption 4 implies that $\Psi_{i,k}(\tau(t-s))$ satisfies all hypothesis of lemma 10, and as consequence we have that $Z_t^{i,k}$ converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$, which concludes the proof.

■

D. Convergence of $\varphi_t^{1,1}$

After studying the asymptotic behavior of the martingale B_t and the rescaled processes $\varphi_t^{i,k}$ for $(i,k) \neq (1,1)$, we study the asymptotic behavior of $\varphi_t^{1,1}$. We start by rewriting it in a more convenient form, using Eqn. (15):

$$\varphi_t^{1,1} = \tilde{\mu}^{1,1}(\frac{1}{\tau} + \int_0^t \Psi_{1,1}(\tau(t-s))ds) \tag{18}$$
$$+ \int_0^t \nu_1 D_{1,1} e^{-\nu_1 D_{1,1}\lambda(t-s)}\sqrt{\pi\varphi_s^{1,1}} dB_s^{1,1} + U_t + V_t, \tag{19}$$

where $\pi = (\sum_i v_{1,i}^2 (v^{-1})_{i,1})(\sum_k \rho_{k,1}^2(\rho^{-1})_{1,k})$,

$$U_t = \int_0^t \Psi_{1,1}(\tau(t-s))(\sqrt{(v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2}} - \sqrt{\pi\varphi_s^{1,1}})dB_s^{1,1}. \tag{20}$$

and

$$V_t = \int_0^t \left(\Psi_{1,1}(\tau(t-s)) - \nu_1 D_{1,1} e^{-\nu_1 D_{1,1}\lambda(t-s)}\right)\sqrt{\pi\varphi_s^{1,1}}dB_s^{1,1} \tag{21}$$

We begin by studying the asymptotic behavior of U_t in Eqn. (20) and V_t in (21).

Lemma 6. *We have that U_t defined in Eqn. (20) converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.*

Proof: Let us define the martingale $Z_t = \int_0^t \Psi_{1,1}(-\tau s)(\sqrt{(v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2}} - \sqrt{\pi\varphi_s^{1,1}})dB_s^{1,1}$, such that $U_t = \Psi_{1,1}(\tau t)Z_t$. Using the product formula for semimartingales and the fact that $\Psi_{1,1}$ has bounded variation, we have that

$$U_t = \int_0^t \partial_t\Psi_{1,1}(\tau s)Z_s ds + \int_0^t \Psi_{1,1}(\tau s)dZ_s = V_t + W_t,$$

where $V_t = \int_0^t \partial_t\Psi_{1,1}(\tau s)Z_s ds$ has bounded variation and $W_t = \int_0^t \Psi_{1,1}(\tau s)dZ_s$ is a martingale with quadratic variation $[W]_t$ satisfying by lemma 8

$$\begin{aligned}
[W]_t &= \int_0^t \Psi_{1,1}^2(\tau s)d[Z]_s \\
&= \int_0^t \Psi_{1,1}^2(\tau s)\Psi_{1,1}^2(-\tau s) \\
&\quad (\sqrt{(v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2}} - \sqrt{\pi\varphi_s^{1,1}})^2 d[B^{1,1}]_s \\
&= \int_0^t (\sqrt{(v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2}} - \sqrt{\pi\varphi_s^{1,1}})^2 d[B^{1,1}]_s.
\end{aligned}$$

Thus, using the fact that $\sqrt{a+b} - \sqrt{b} \leq \frac{a}{2\sqrt{b}}$ for $a, b > 0$, we have by lemma 4

$$\begin{aligned}
\mathbb{E}[[W]_t] &= \mathbb{E}[\int_0^t (\sqrt{(v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2}} - \sqrt{\pi\varphi_s^{1,1}})^2 ds] \\
&\leq \mathbb{E}[\int_0^t \frac{((v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2} - \pi\varphi_s^{1,1})^2}{4\pi\varphi_s^{1,1}} ds] \\
&\leq L^2\mathbb{E}[\int_0^t \frac{(\sum_{(i,k)\neq(1,1)}\varphi_s^{i,k})^2}{4\pi\varphi_s^{1,1}} ds]
\end{aligned}$$

for some constant $L > 0$ by lemma 11. Since $\varphi_s^{1,1} \geq \frac{\tilde{\mu}^{1,1}}{\tau} \geq \frac{c}{\tau} > 0$, we have by lemma 4 that for $t \in [0,1]$

$$\begin{aligned}
\mathbb{E}[[W]_t] &\leq \frac{\tau L^2}{4\pi c}\mathbb{E}[\int_0^t (\sum_{(i,k)\neq(1,1)} \varphi_s^{i,k})^2 ds] \\
&\leq \frac{\tau L(NK-1)}{4\pi c}\mathbb{E}[\int_0^t \sum_{(i,k)\neq(1,1)} (\varphi_s^{i,k})^2 ds] \\
&= \frac{\tau L(NK-1)}{4\pi c}\int_0^t \sum_{(i,k)\neq(1,1)} \mathbb{E}[(\varphi_s^{i,k})^2]ds \\
&\leq \frac{\tau L^2(NK-1)^2}{4\pi c}\int_0^t \frac{1}{\tau^2}ds = \frac{L't}{\tau} \leq \frac{L'}{\tau}
\end{aligned}$$

for $L' = \frac{L^2(NK-1)^2}{4\pi c}$.

Thus, by Markov's inequality we have that for all $\varepsilon > 0$ and for all $t \in [0,1]$

$$\mathbb{P}([W]_t \geq \varepsilon) \leq \frac{L'}{\varepsilon}(\frac{1}{\tau^2} + \frac{1}{\tau}) \to 0 \quad \text{when} \quad \tau \to \infty,$$

which proves that $[W]_t$ converges in probability to 0 for all $t \geq 0$.

Since W has uniformly bounded jumps, we have by theorem $VIII.3.11$ of [35] that W_t converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.

Now, regarding V_t, we have that since $|\partial_t\Psi_{1,1}(\tau t)| \leq C$ for some constant $C > 0$,

$$\begin{aligned}
\mathbb{E}[(V_t - V_s)^2] &= \mathbb{E}[(\int_s^t \partial_t\Psi_{1,1}(\tau u)Z_u du)^2] \\
&\leq C^2(t-s)^2\mathbb{E}[(\sup_{u\in[s,t]} Z_u)^2] \leq C'(t-s)^2\mathbb{E}[[Z]_t]
\end{aligned}$$

by the Burkholder-Davis-Gundy inequality. Since by lemma 8

$$[Z]_t = \int_0^t \Psi_{1,1}^2(-\tau s)(\sqrt{(v_1^{\odot 2})^T v^{-1}\varphi_s\rho^{-1}\rho_1^{\odot 2}} - \sqrt{\pi\varphi_s^{1,1}})^2 d[B^{1,1}]_s$$

and $\Psi_{1,1}(-\tau s) \leq C$ for some constant $C > 0$, we have using the same calculations as before and choosing $s = 0$ that for $t \in [0,1]$

$$\mathbb{E}[V_t^2] \leq C't^2\mathbb{E}[[Z]_t] \leq \frac{C''}{\tau^2},$$

which easily implies that $(V_{t_1}, \cdots, V_{t_n}) \to 0$ in distribution for every $(t_1, \cdots, t_n) \in [0,1]^n$ when $\tau \to \infty$, i.e., we have the convergence of the finite-dimensional distribution of V_t to 0 when $\tau \to \infty$.

Moreover, since $\mathbb{E}[(V_t - V_s)^2] \leq C'''(t-s)^2$, we have by the Kolmogorov criterion for tightness that V_t is tight for the Skorohod topology in $[0,1]$, which implies that V_t converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.

Hence, we clearly have that $U_t = V_t + W_t$ converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$. ∎

Lemma 7. *We have that V_t defined in Eqn. (21) converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.*

Proof: Define the function

$$\begin{aligned}
f_\tau(t) &= \Psi_{1,1}(\tau t) - \nu_1 D_{1,1}e^{-\nu_1 D_{1,1}\lambda t} \\
&= \nu_1 D_{1,1}\left(e^{-\omega_\tau\tau(1-\frac{\nu_1 D_{1,1}}{\omega_\tau})t} - e^{-\nu_1 D_{1,1}\lambda t}\right).
\end{aligned}$$

By assumption 4, that there exists a $C > 0$ such that

1) $\sup_\tau \sup_t |f_\tau(t)| \leq C$,
2) Since f_τ is a difference of exponential functions, we can assume without loss of generality that $|\hat{f}_\tau(z)| \leq C(|\frac{1}{z}| \wedge 1)$,
3) Applying lemma 4.7 of [20] we have that for any $0 < \varepsilon < 1$, there exists $C_\varepsilon > 0$ such that for every t, s

$$\sup_\tau \int_{\mathbb{R}} (f_\tau(t-u) - f_\tau(s-u))^2 du \leq C_\varepsilon|t-s|^{1-\varepsilon},$$

4) Since

$$f_\tau^2(t) = \nu_1^2 D_{1,1}^2 \Big(\Psi_{1,1}^2(\tau t) + e^{-2\nu_1 D_{1,1}\lambda t} - 2\Psi_{1,1}(\tau t) e^{-\nu_1 D_{1,1}\lambda t} \Big),$$

we have that

$$\int_{\mathbb{R}^+} f_\tau^2(t)dt = \nu_1^2 D_{1,1}^2 \Big(\frac{1}{2\omega_\tau \tau (1 - \frac{\nu_1 D_{1,1}}{\omega_\tau})} + \frac{1}{2\nu_1 D_{1,1}\lambda} - 2\frac{1}{\omega_\tau \tau (1 - \frac{\nu_1 D_{1,1}}{\omega_\tau}) + \nu_1 D_{1,1}\lambda} \Big) \to 0.$$

5) Since, for $\alpha > 0$, $e^{-\alpha t}$ satisfies $|e^{-\alpha t} - e^{-\alpha s}| \leq \alpha|t-s|$, we easily have that there exists a constant $C > 0$ such that

$$|f_\tau(t) - f_\tau(s)| \leq C\tau|t-s|.$$

Hence, f_τ satisfies all hypothesis of lemma 10. Moreover, $g(\varphi_s) = \sqrt{\pi \varphi_s^{1,1}}$ easily satisfies

$$|g(\varphi_t)| \leq C(1 + ||\varphi_t||).$$

We can thus apply lemma 10 to conclude the proof. ∎

We have arrived to the final step of the proof: by lemma 2, we have that $\tilde{\mu}^{1,1} \int_0^t \Psi_{1,1}(\tau(t-s))ds$ converges uniformly in $[0,1]$ to $\tilde{\mu}^{1,1} \int_0^t \nu_1 D_{1,1} e^{-\nu_1 D_{1,1}\lambda(t-s)}ds = \tilde{\mu}^{1,1}(\frac{1-e^{-\nu_1 D_{1,1}\lambda t}}{\lambda})$, when $\tau \to \infty$.

Moreover, by lemma 5 we have that $B_t^{1,1}$ converges in law to a standard Brownian motion for the Skorohod topology in $[0,1]$ when $\tau \to \infty$, and by lemmas 6 and 7 we have that U_t and V_t converge in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.

As in [20], since U_t and V_t converge to a deterministic limit, we get the convergence in law, for the product topology, of the triple $(U_t, V_t, B_t^{1,1})$ to $(0, 0, W_t)$ with W a standard Brownian motion. The components of $(0, 0, W_t)$ being continuous, the last convergence also takes place for the Skorohod topology on the product space.

Thus, we have by theorem 5.4 of [34] that $\varphi_t^{1,1}$ converges in law to the limit process C_t for the Skorohod topology in $[0,1]$ when $\tau \to \infty$, where C_t is the unique solution of

$$C_t = \tilde{\mu}^{1,1}(\frac{1 - e^{-\nu_1 D_{1,1}\lambda t}}{\lambda}) + \nu_1 D_{1,1} \int_0^t e^{-\nu_1 D_{1,1}\lambda(t-s)} \sqrt{\pi C_s} dW_s,$$

where W_t is a standard Brownian motion.

By a simple calculation, we have that C_t satisfies the following stochastic differential equation

$$\begin{aligned} dC_t &= \nu_1 D_{1,1}\tilde{\mu}^{1,1} e^{-\nu_1 D_{1,1}\lambda t}dt \\ &+ \nu_1 D_{1,1}\Big(-\nu_1 D_{1,1}\lambda \int_0^t e^{-\nu_1 D_{1,1}\lambda(t-s)}\sqrt{\pi C_s}dW_s dt \\ &+ \sqrt{\pi C_t}dW_t \Big) \\ &= \nu_1 D_{1,1}\tilde{\mu}^{1,1} e^{-\nu_1 D_{1,1}\lambda t}dt \\ &+ \nu_1 D_{1,1}\Big((\tilde{\mu}^{1,1}(1 - e^{-\nu_1 D_{1,1}\lambda t}) - \lambda C_t)dt + \sqrt{\pi C_t}dW_t \Big) \\ &= \nu_1 D_{1,1}\lambda(\frac{\tilde{\mu}^{1,1}}{\lambda} - C_t)dt + \nu_1 D_{1,1}\sqrt{\pi C_t}dW_t. \end{aligned}$$

Remark: One promptly has that the columns of v^{-1} are the right-eigenvectors of J and that the rows of ρ^{-1} are the left-eigenvectors of B, thus $\pi > 0$ can be rewritten as

$$\pi = (\sum_i v_{1,i}^2 \tilde{v}_{i,1})(\sum_k \rho_{k,1}^2 \tilde{\rho}_{1,k}),$$

where v_1^T is the leading left-eigenvector of J, $\tilde{v}_1$ is the right-eigenvector of J, ρ_1 is the leading right-eigenvector of B and $\tilde{\rho}_1^T$ is the leading left-eigenvector of B. Moreover, by the Perron-Frobenius theorem we have that v, $\tilde{v}$, ρ and $\tilde{\rho}$ have nonnegative entries.

Remark: In the one-dimensional case, we clearly have that $\pi = 1$, retrieving thus the same result as in [20].

APPENDIX C
ADDITIONAL LEMMAS

Lemma 8. *Let $f : \mathbb{M}_{N\times K}(\mathbb{R}^+) \to \mathbb{R}^+$ and $g : \mathbb{M}_{N\times K}(\mathbb{R}^+) \to \mathbb{R}^+$ be functions satisfying for some constant $C > 0$*

$$|f(\varphi_t)| \leq C(1 + ||\varphi_t||) \ \ \textit{and} \ \ |g(\varphi_t)| \leq C(1 + ||\varphi_t||),$$

let $h : \mathbb{R} \to \mathbb{R}$ and $r : \mathbb{R} \to \mathbb{R}$ be continuous functions and let Z_t^1 and Z_t^2 be L^2 martingales such that $[Z^1, Z^2]_t = t + M_t$, where M_t is a martingale.

Defining $z_t^1 = \int_0^t h(s)f(\varphi_s)dZ_s^1$ and $z_t^2 = \int_0^t r(s)g(\varphi_s)dZ_s^2$ we have that

$$\mathbb{E}[z_t^1 z_t^2] = \int_0^t h(s)r(s)\mathbb{E}[f(\varphi_s)g(\varphi_s)]ds.$$

Moreover, if Z_t is a L^2 semimartingale, we have that the stochastic process $Y_t = \int_0^t h(s)f(\varphi_s)dZ_s$ satisfies

$$[Y]_t = \int_0^t h^2(s)f^2(\varphi_s)d[Z]_s \ \ \textit{and} \ \ \mathbb{E}[Y_t^2] \leq \mathbb{E}[[Y]_t].$$

Lemma 9. *Let X be a $N \times K$ matrix with nonnegative entries, $v^T \neq 0$ be a $1 \times N$ row vector with nonnegative entries and $\rho \neq 0$ be a $K \times 1$ vector with nonnegative entries. Then*

$$(v^{\odot 2})^T X \rho^{\odot 2} \leq ||v||.||\rho||.v^T X\rho.$$

Proof: Define the row vector $\tilde{v}^T = \frac{v^T}{||v||}$ and the vector $\tilde{\rho} = \frac{\rho}{||\rho||}$, such that $\tilde{v}_i^T \leq 1$ and $\tilde{\rho}_k \leq 1$, which implies $(\tilde{v}_i)^2 \leq \tilde{v}_i$ and $(\tilde{\rho}_k)^2 \leq \tilde{\rho}_k$. Then

$$\begin{aligned}(v^{\odot 2})^T X \rho^{\odot 2} &= ||v||^2.||\rho||^2 \sum_{i,k} \tilde{v}_i^2 X_{i,k} \tilde{\rho}_k^2 \\ &\leq ||v||^2.||\rho||^2. \sum_{i,k} \tilde{v}_i X_{i,k} \tilde{\rho}_k \\ &= ||v||.||\rho||. \sum_{i,k} v_i X_{i,k} \rho_k = ||v||.||\rho|| v^T X \rho.\end{aligned}$$

■

This lemma can be proven using the ideas in [20] (see the proof of the convergence for the rescaled process $(Y_t^T)_{t\in[0,1]}$ at the beginning of page 18, corollaries 4.1, 4.2, 4.3, 4.4 and lemma 4.7).

Lemma 10. *Let $f_\tau : \mathbb{R}^+ \to \mathbb{R}$ be a sequence of functions such that*

1) *There exists a constant $C > 0$ such that $\sup_\tau \sup_{t\in\mathbb{R}} |f_\tau(t)| \leq C$,*
2) *There exists a constant $C > 0$ such that for all τ*

$$|f_\tau(t) - f_\tau(s)| \leq C\tau|t-s|,$$

3) *For any $0 < \varepsilon < 1$, there exists $C_\varepsilon > 0$ such that for every t, s*

$$\sup_\tau \int_{\mathbb{R}} (f_\tau(t-u) - f_\tau(s-u))^2 du \leq C_\varepsilon |t-s|^{1-\varepsilon},$$

4) *$\int_{\mathbb{R}^+} f_\tau^2(s)ds \to 0$ when $\tau \to \infty$, and*
5) *There exists a constant $C > 0$ such that $\sup_\tau |\hat{f}_\tau(z)| \leq C(|\frac{1}{z}| \wedge 1)$.*

Let $g : \mathcal{M}_{N\times K}(\mathbb{R}^+) \to \mathbb{R}$ be a function satisfying for some constant $C > 0$

$$|g(\varphi_t)| \leq C(1 + ||\varphi_t||),$$

and define $Y_t^{i,k,\tau} = \int_0^t f_\tau(t-s) g(\varphi_s) dB_s^{i,k}$.

We have that $Y_t^{i,k,\tau}$ converges in law to 0 for the Skorohod topology in $[0,1]$ when $\tau \to \infty$.

Lemma 11. *Let $\pi = (\sum_i v_{1,i}^2 (v^{-1})_{i,1})(\sum_k \rho_{k,1}^2 (\rho^{-1})_{1,k})$. We have that*

$$(v_1^{\odot 2})^T v^{-1} \varphi_t \rho^{-1} \rho_1^{\odot 2} \leq \pi \varphi_t^{1,1} + L \sum_{(i,k)\neq(1,1)} \varphi_t^{i,k}$$

for some constant $L > 0$.

Proof: Let us define the $1 \times N$ row vector $V^T = (v_1^{\odot 2})^T v^{-1}$ and the $K \times 1$ vector $R = \rho^{-1} \rho_1^{\odot 2}$, such that

$$V_j^T = \sum_i v_{1,i}^2 v_{i,j}^{-1} \quad \text{and} \quad R_c = \sum_k \rho_{c,k}^{-1} \rho_{k,1}^2.$$

Thus

$$\begin{aligned}(v_1^{\odot 2})^T v^{-1} \varphi_t \rho^{-1} \rho_1^{\odot 2} &= \sum_{j,c} V_j^T \varphi_t^{j,c} R_c \\ &= \pi \varphi_t^{1,1} + \sum_{(j,c)\neq(1,1)} V_j^T \varphi_t^{j,c} R_c \\ &\leq \pi \varphi_t^{1,1} + L \sum_{(i,k)\neq(1,1)} \varphi_t^{i,k}\end{aligned}$$

for a constant $L > 0$. ■

References

[1] J. Kleinberg, “Bursty and hierarchical structure in streams,” *In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 91–101, 2002.

[2] X. Wang, C. Zhai, and R. S. X. Hu, “Mining correlated bursty topic patterns from coordinated text streams,” *In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007.

[3] M. Cataldi, L. D. Caro, , and C. Schifanella, “Emerging topic detection on twitter based on temporal and social terms evaluation,” *In Proceeding of the 10th International Workshop on Multimedia Data Mining (MDMKDD)*, 2010.

[4] T. Takahashi, R. Tomioka, and K. Yamanishi, “Discovering emerging topics in social streams via link anomaly detection,” *In Proceeding of the 11th IEEE International Conference on Data Mining (ICDM)*, 2011.

[5] J. Goldenberg, B. Libai, and E. Muller, “Using complex systems analysis to advance marketing theory development,” *Academy of Marketing Science Review*, 2001.

[6] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” *In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146, 2003.

[7] H. P. Young, “The dynamics of social innovation,” *Proc Natl Acad Sci USA*, vol. 108, no. Suppl. 4, pp. 21 285–21 291, 2011.

[8] Y. Altshuler, W. Pan, and A. Pentland, “Trends prediction using social diffusion models,” *Social Computing, Behavioral - Cultural Modeling and Prediction*, vol. 7227, pp. 97–104, 2008.

[9] R. Crane and D. Sornette, “Robust dynamic classes revealed by measuring the response function of a social system,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15 649–15 653, 2008.

[10] L. Li and H. Zha, “Dyadic event attribution in social networks with mixtures of Hawkes processes,” *Proceedings of 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2013.

[11] S.-H. Yang and H. Zha, “Mixture of mutually exciting processes for viral diffusion,” *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.

[12] I. Valera, M. Gomez-Rodriguez, and K. Gummadi, “Modeling adoption and usage frequency of competing products and conventions in social media,” *Workshop in Networks: From Graphs to Rich Dat at Neural Information Processing Systems Conference (NIPS)*, 2014.

[13] M. Farajtabar, N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song, “Shaping social activity by incentivizing users,” *In proceedings of Neural Information Processing Systems Conference (NIPS)*, 2014.

[14] J. Cheng, L. Adamic, A. Dow, J. Kleinberg, and J. Leskovec, “Can cascades be predicted?” *In Proceedings of ACM International Conference on World Wide Web (WWW)*, 2014.

[15] J. Leskovec, L. Backstrom, and J. Kleinberg, “Meme-tracking and the dynamics of the news cycle,” *In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 497–506, 2009.

[16] R. Dawkins, *The Selfish Gene*, 2nd ed. Oxford University Press, 1989.

[17] A. G. Hawkes, “Spectra of some self-exciting and mutually exciting point processes,” *Biometrika*, vol. 58, pp. 83–90, 1971.

[18] T. Liniger, “Multivariate Hawkes processes,” *ETH Doctoral Dissertation*, no. 18403, 2009.

[19] M. E. J. Newman, “The mathematics of networks,” *Notes*, 2006.

[20] T. Jaisson and M. Rosenbaum, “Limit theorems for nearly unstable hawkes processes,” *ArXiv: 1310.2033*, 2013.

[21] G.-E. Espinosa and N. Touzi, “Detecting the maximum of a scalar diffusion with negative drift,” *SIAM Journal on Control and Optimization*, vol. 50, no. 5, pp. 2543–2572, 2012.

[22] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, ser. Springer series in Statistics. Springer, 2005.

[23] Y. Ogata, “On lewis simulation method for point processes,” *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23–31, 1981.

[24] A. G. Hawkes and D. Oakes, “A cluster process representation of a self-exciting point process,” *J. Appl. Prob.*, vol. 11, pp. 493–503, 1974.

[25] P. Brémaud and L. Massoulié, “Stability of nonlinear Hawkes processes,” *The Annals of Probability*, vol. 24, no. 3, pp. 1563–1588, 1996.

[26] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,” *Journal of Complex Networks*, no. 2, 2014.

[27] P. Holme and J. Saramäki, (eds.), *Temporal Networks*. Berlin: Springer, 2013.

[28] E. Bacry and J.-F. M. S. Gaïffas, “Concentration for matrix martingales in continuous time and microscopic activity of social networks,” *ArXiv: 1412.7705*, 2014.

[29] J. C. Cox, J. E. Ingersoll, and S. A. Ross, “A theory of the term structure of interest rates,” *Econometrica*, vol. 53, pp. 385–407, 1985.

[30] D. F. Gleich, “Pagerank beyond the web,” *ArXiv: 1407.5107*, 2014.

[31] P. Billingsley, *Convergence of probability measures*, ser. Wiley series in Probability and Statistics. New York: Wiley, 2009, vol. 493.

[32] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1974.

[33] A. Kuznetsov, “Solvable markov processes,” Ph.D. dissertation, University of Toronto, 2004.

[34] T. G. Kurtz and P. Protter, “Weak limit theorems for stochastic integrals and stochastic differential equations,” *The Annals of Probability*, vol. 19, no. 3, pp. 1035–1070, 1991.

[35] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*. Berlin: Springer-Verlag, 1987, vol. 288.