

Trend detection in social networks using Hawkes processes

Julio Cesar Louzada Pinto

Institut Mines-Telecom
Telecom SudParis

UMR CNRS 5157, France

Email: julio.louzada_pinto@telecom-sudparis.eu

Tijani Chahed

Institut Mines-Telecom
Telecom SudParis

UMR CNRS 5157, France

Email: tijani.chahed@telecom-sudparis.eu

Eitan Altman

INRIA Sophia Antipolis Méditerranée
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 to 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 data necessary to the detection.

I. INTRODUCTION

We introduce in this paper 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 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.

In order to fully exploit the social ties between users and information in social networks, we base our trend detection algorithm on information diffusion models [5], [6], [7], [8], and more specifically on a Hawkes-based model for information diffusion in social networks [9], [10], [11], [12]. The Hawkes-based model allows: 1) leveraging on the knowledge of the influences between users and contents, 2) to fully explore the real time of broadcasts, 3) leveraging on the knowledge of users intrinsic (or exogenous) rates. Moreover, the Hawkes 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].

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 user 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 [14]. 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.

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 have a 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 against their expected maximum values at each time, meaning that topics that do not have yet large intensities can indeed be trendy. Still, our algorithm is also able capturing the trendiness coming from large intensities.

The proposed method bares some resemblance with classical works on trend detection. For example, as mentioned above, our algorithm uses Hawkes processes [15], [16] to model the broadcasting/posting times of messages in a social network, which is similar to the infinite-state automaton approach of Kleinberg [1]; the difference between both approaches is how to deal with the intensity stemming from the broadcasting activity: while Kleinberg searches the periods in time with a high frequency of broadcasts about similar contents, we study a Hawkes intensity for broadcasts about contents that can increase even by broadcasted messages about different ones. Since the influences of users and topics in our information diffusion Hawkes model generates correlation in broadcasts between different contents, our work also relates through the underlying Hawkes intensity to the work of Wang *et al.* [2], where the authors propose a probabilistic algorithm that discovers correlated bursty patterns and their periods across text streams; the main difference besides the underlying information diffusion model is that we assume the broadcasts to be about specific predefined topics, whereas Wang *et al.* use text mining techniques to unravel the topics, defined as probabilities over vocabularies.

In comparison to other works on trend detection in social networks, our framework resembles the one proposed by Cataldi *et al.* [3], where the authors devise an algorithm to detect real-time emerging topics in Twitter firstly by extracting

the contents of the tweets with a model for their life cycle and secondly by considering the social importance of the sources of the tweets, using the Page Rank algorithm. It also resembles the one proposed by Takahashi *et al.* [4], where the authors derive an algorithm focusing on the social aspects of social networks by dynamically generated links between users, and propose a stochastic model for behavior of a generic social network user, detecting the emergence of a new topic. Again, the major difference between these works and ours is the underlying model of broadcasts and the fact that our methodology does not rely on text mining techniques since the content of each broadcasted message is assumed to be already labeled.

To the best of our knowledge, our proposal 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 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 [15], [16].

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 [17], and as such they are designed to model discrete events in networks such as posting, sharing, tweeting, liking, 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 [17], [18], which facilitates a maximum likelihood estimation of the parameters.
- A linear Hawkes process can be seen as a Poisson cluster process [19], which permits the distinction of two regimes: a stationary (or stable) regime in which the intensity processes has a stationary version and thus a number of events that is in average linear in

time, and a nonstationary (or unstable) regime, in which the process has a superlinear average number of events in time (see [15], [20] for details).

- It easily allows extensions from the basic model, such as multiple social networks [21], dynamic/temporal networks [22], seasonality and/or time-dependence for the intrinsic diffusion rate of users [12], etc.

We represent our social network as a communication graph $G = (V, E)$, where V is the set of users with cardinality $\#V = N$ and E is the edge set, i.e., the set with all the possible communication links between users, basing ourselves on the information diffusion model described in [23]. 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 predefined¹ topics (for example 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 $t \in [0, \tau]$, where τ is the moment of predicting the trendiest topics.

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

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

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

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 $\phi(t) = e^{-\omega t} \cdot \mathbb{I}_{\{t>0\}}$ a light-tailed exponential kernel [11] and $\phi(t) = (a+t)^{-b} \cdot \mathbb{I}_{\{t>0\}}$ a heavy-tailed power-law kernel [9].

The intensity can be seen in matrix form as

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

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 procedures can be used to estimate J and B , as in [11], [23].

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 [19], which gives us the following result (whose proof is well explained in [15], [20]):

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 condition²*

$$sp(J)sp(B)\|\phi\|_1 < 1. \quad (2)$$

III. DISCOVERING TRENDY TOPICS

After defining in detail 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 , $k \in \{1, 2, \dots, K\}$. It consists of the following steps:

- 1) Perform a temporal rescaling of the intensity following the theory of nearly unstable Hawkes processes [24], which gives a Cox-Ingersoll-Ross (CIR) process [25] as the limiting rescaled process.
- 2) Search the expected maxima of the rescaled intensities for each topic $k \in \{1, 2, \dots, K\}$, with the aid of the limit CIR process. This task is achieved by solving stochastic control problems V_k following the theory developed in [26], 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 $k \in \{1, 2, \dots, K\}$.

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

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 [24].

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 thus the number of broadcasts remains at most linear, and an *unstable* regime where the number of broadcasts increases in a superlinear fashion.

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 satisfied but barely, i.e., $sp(J)sp(B)\|\phi\|_1 = 1 - \frac{\lambda}{\tau}$ for $\lambda > 0$, and where there exists a drastic change in the behavior of the broadcasts - a Hawkes process satisfying this property is called *nearly unstable* [24]. 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_t 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, we perform a rescaling to the intensity λ_t , which converges in law when $\tau \rightarrow \infty$ 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 [26].

Remark: As there are several ways to rescale the intensity λ_t and obtain a nontrivial limit behavior, we have chosen to use the framework of [24] because their rescaling transforms λ_t into a mean-reverting Brownian diffusion, for which there exist detailed studies about finding its expected maxima, such as [26].

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 one 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_t as base measure for the trendiness. Moreover, by the previous paragraph, we must also weight the intensity λ_t 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 [14].

Mathematically speaking, let v^T be the left-eigenvector of the user-user interaction matrix J , related to the leading³ 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 [14], [27]) 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

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

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

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

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 $\tilde{\lambda}_t$.

³This 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 $\tilde{\lambda}_t^k$ 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 , which will be determined by Eqn. (9).

Since the influences $\nu(\phi * d\tilde{X})_t B$ are always nonnegative in Eqn. (3), we can only find peaks when $\tilde{\lambda}_t^k$ 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 assumption⁴:

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., $\phi(t) = e^{-\omega t} \mathbb{I}_{\{t \geq 0\}}$, 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

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

⁴The 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 [26]) has not yet been solved analytically, hence numerical methods should be employed.

which when multiplied by P^T by the left becomes

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

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

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

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

Now, following [24], we rescale φ_t by "pushing" the timescale parameter ω to the unstable regime of \tilde{X}_t , so as to obtain a nontrivial behavior (peak) for the intensity $\tilde{\lambda}_t$, 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 $\tau \rightarrow \infty$, which implies (we assume without loss of generality that $\tau > \lambda$)

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

The rescaling stems from the next theorem⁵ (the one-dimensional case is proven in theorem 2.2 of [24]):

Theorem 1. *Let assumption 1 be true, the temporal kernel be defined as $\phi(t) = e^{-\omega t} \mathbb{I}_{\{t>0\}}$, let $\rho = ((P^{-1})_{1,1}, \dots, (P^{-1})_{1,K})$ 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 $\tau \rightarrow \infty$, then the rescaled process $\frac{1}{\tau} \varphi_{\tau t}^1$ converges in law, for the Skorohod⁶ topology in $[0, 1]$, to a CIR process C^1 satisfying the following stochastic differential equation (SDE)

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

where W_t is a standard Brownian motion.

Moreover for $k > 1$, the rescaled processes $\frac{1}{\tau} \varphi_{\tau t}^k$ converge in law to 0, for the Skorohod topology to 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 $\bar{\mu} = \frac{\vartheta^1}{\lambda}$. As already discussed in subsection III-C, if one wants to capture some trend behavior one must see this process above its stationary expectation $\bar{\mu}$, i.e., one must study the process $C_t = C_t^1 - \bar{\mu}$.

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

$$dC_t = -\lambda \nu D_{1,1} C_t dt + \nu D_{1,1} \sqrt{\pi} \sqrt{C_t + \bar{\mu}} dW_t. \quad (8)$$

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 $\tilde{\lambda}_t$ using the framework developed by Espinosa and Touzi [26] 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 $\tilde{\lambda}_t^k$ from its peak, where a peak is represented by a maximum of $\tilde{\lambda}_t^k$ in the sense of [26]. To do so, we use the fact that $\tilde{\lambda}_t^k = (P^{-1})^T \varphi_t$ to determine the limit behavior of $\frac{\tilde{\lambda}_{\tau t}^k}{\tau}$, namely $\tilde{\lambda}_{t,\infty}^k$, as

$$\tilde{\lambda}_{t,\infty}^k = \sum_j P_{j,k}^{-1} C_t^j = P_{1,k}^{-1} C_t^1 = P_{1,k}^{-1} (C_t + \bar{\mu}),$$

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

$$V_k = \inf_{\theta \in \mathcal{T}_0} \mathbb{E} \left[\frac{(P_{1,k}^{-1})^2}{2} (C_{T_0}^* - C_\theta)^2 \right], \quad (9)$$

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

By the theory developed in [26], 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 $\gamma_{\downarrow}^k(x)$ and then an increasing part $\gamma_{\uparrow}^k(x)$, which are found by solving the ordinary differential equations (ODE) (5.1) and (5.15) in [26], respectively⁷.

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

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

⁷For the CIR case we have by Eqn. (8) that the functions α , S and S' defined in [26] are

- $\alpha(x) = \frac{2\lambda x}{\nu D_{1,1} \pi (x + \bar{\mu})}$, $S'(x) = e^{\frac{2\lambda x}{\nu D_{1,1} \pi} (\frac{x}{\bar{\mu}} + 1)} - \frac{2\lambda \bar{\mu}}{\nu D_{1,1} \pi}$ 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 [29]), since it must satisfy $S(0) = 0$ and $S'(0) = 1$ (see [30]).

⁵For a proof of this theorem, please check the full paper, reference [d11] at <http://www-sop.inria.fr/members/Eitan.Altman/dodescaden.html>.

⁶The 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 [28] for a formal definition.

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

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

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

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}, \dots, P_{K1}) , left-eigenvector $(P_{11}^{-1}, \dots, P_{1K}^{-1})$ 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 $\{0, \delta, 2\delta, \dots, (T-1)\delta\}$, following [26]
 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

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 $\bar{\mu}$, which has complexity $\mathcal{O}(NK)$.

IV. NUMERICAL EXAMPLES

We provide in this section two examples where we apply our trend detection algorithm.

The first example is performed on a synthetic near unstable Hawkes processes in a social network using Ogata's thinning algorithm⁸ [18] in a time horizon $\tau = 50$. We use 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 and 2.

The second example is applied to a MemeTracker dataset containing different memes (short distinct phrases) for the 5,000 most active sites from 4 million sites from March 2011 to February 2012⁹. We use the 10 most broadcasted memes, which are: 1) dancing with the stars, 2) two and a half men, 3) sex and the city, 4) rolling in the deep, 5) too big to fail, 6) don't ask, don't tell, 7) i have a dream, 8) i will always love you, 9) the girl with the dragon tattoo, 10) the tree of life.

⁸The 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 $\frac{M - \sum_{i,k} \lambda_t^{i,k}}{M}$.

⁹Data available at <http://snap.stanford.edu/infopath>.

In this numerical example, each meme plays the role of a topic in our theoretical model of section II. We use a maximum likelihood estimation procedure for the parameters (see [23]).

For both examples, we shall illustrate how our method is able to detect the trendiness of each topic or meme according to the index \mathcal{I} , and that the highest trendiness does not necessarily correspond to the topic or meme which has the highest number of broadcasts.

For both examples, figures 1 and 3 plot the scaled topic intensities $\frac{\tilde{\lambda}_{\tau,t}}{\tau}$ as a function of time, and figures 2 and 4 plot the cumulative number of broadcasts about each topic as a function of time, i.e., $\bar{X}_t^k = \sum_i X_t^{i,k}$. Furthermore, we compute in tables I and II the trend indices $\tilde{\mathcal{I}}^k$ and the total number of broadcasts for both examples. We use 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 I, one can see that the trend index for topic 1 is the highest, even though it does not possess the highest number of broadcasts, which is held by topic 3. The reason is that in the synthetic dataset, topic 1 has the largest topic intensity.

In reference to table II, one can see that meme 9 shows higher trendiness than the other memes, even though it does not possess the highest intensity and it possesses the smallest total number of broadcasts; it is then followed by memes 10 and 8 in second and third place, respectively. The reason is similar for all of them: they possess larger and more frequent "peaks" of intensity compared to other memes, occurring at times closer to the prediction instant τ , as depicted in figure 3.

A different phenomenon occurs for meme 2, which has the highest total number of broadcasts but the least trendiness. Since most of the broadcasts of meme 2 occur very early in time, the peak of intensity related to this increase in the number of broadcasts has little impact in the trend index, which takes more into account broadcasts that happen near the prediction instant τ . Thus, as meme 2 does not have significant peaks in intensity near the prediction instant τ , it receives a lower trend index. *Both phenomena illustrate the difference between our algorithm and one that looks solely to the largest topic intensities and the total number of broadcasts.*

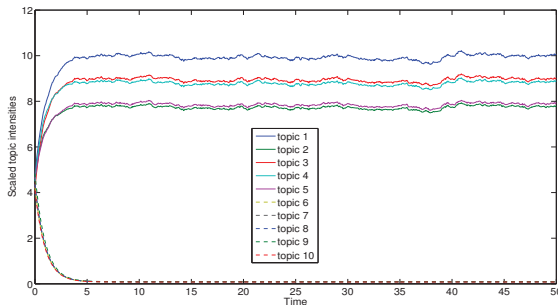


Fig. 1. Topic intensities $\tilde{\lambda}_t^k = \sum_i \lambda_t^{i,k} v_i$ for the synthetic dataset.

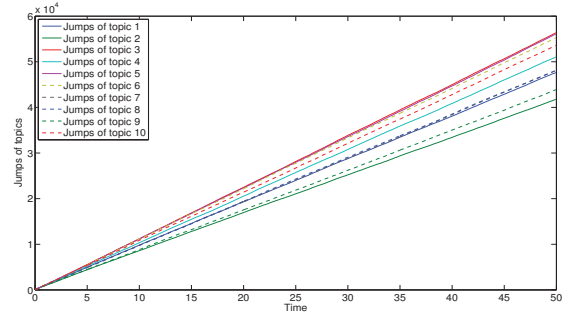


Fig. 2. Cumulative sum of jumps of topics $\bar{X}_t^k = \sum_i X_t^{i,k}$ for the synthetic dataset.

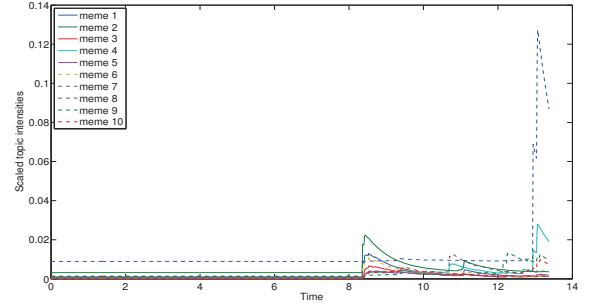


Fig. 3. Topic intensities $\tilde{\lambda}_t^k = \sum_i \lambda_t^{i,k} v_i$ for the meme tracker dataset.

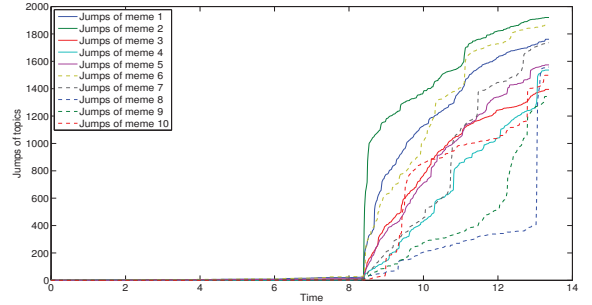


Fig. 4. Cumulative sum of jumps of topics $\bar{X}_t^k = \sum_i X_t^{i,k}$ for meme tracker dataset.

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 messages in the social network can be modeled 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 defined our idea of trendiness and derived trend indices for each topic being disseminated. These indices take into consideration the time between the prediction of the trendy topics and the broadcast times of messages, the intensity of broadcasting messages, 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,

TABLE I. COMPARISON OF INDICES FOR SYNTHETIC DATASET.

	TOPIC 1	TOPIC 2	TOPIC 3	TOPIC 4	TOPIC 5	TOPIC 6	TOPIC 7	TOPIC 8	TOPIC 9	TOPIC 10
\tilde{I}	0.1038	0.0880	0.0968	0.0955	0.08891	0.0334	0.0334	0.0334	0.0334	0.0334
NUMBER OF POSTS	47640	41770	56368	51039	56097	55252	48105	48096	43882	53580

TABLE II. COMPARISON OF INDICES FOR MEMETRACKER DATASET.

	MEME 1	MEME 2	MEME 3	MEME 4	MEME 5	MEME 6	MEME 7	MEME 8	MEME 9	MEME 10
\tilde{I}	0.0023	0.0014	0.0042	0.0087	0.0045	0.0033	0.0071	0.0090	0.01590	0.0106
NUMBER OF POSTS	1768	1925	1406	1537	1578	1871	1746	1562	1344	1499

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 a suitable stochastic control problem. 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.

ACKNOWLEDGMENT

This work was jointly supported by Institut Mines Telecom Future et Ruptures program and the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672.

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 Data" 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] M. E. J. Newman, "The mathematics of networks," *Notes*, 2006.
- [15] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, pp. 83–90, 1971.
- [16] T. Liniger, "Multivariate Hawkes processes," *ETH Doctoral Dissertation*, no. 18403, 2009.
- [17] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, ser. Springer series in Statistics. Springer, 2005.
- [18] Y. Ogata, "On lewis simulation method for point processes," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23–31, 1981.
- [19] 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.
- [20] P. Brémaud and L. Massoulié, "Stability of nonlinear Hawkes processes," *The Annals of Probability*, vol. 24, no. 3, pp. 1563–1588, 1996.
- [21] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, no. 2, 2014.
- [22] P. Holme and J. Saramäki, (eds.), *Temporal Networks*. Berlin: Springer, 2013.
- [23] J. C. Louzada Pinto and T. Chahed, "Modeling user and topic interactions in social networks using hawkes processes," *8th International Conference on Performance Evaluation Methodologies and Tools (VAL-UETOOLS '14)*, 2014.
- [24] T. Jaisson and M. Rosenbaum, "Limit theorems for nearly unstable hawkes processes," *ArXiv: 1310.2033*, 2013.
- [25] 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.
- [26] 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.
- [27] D. F. Gleich, "Pagerank beyond the web," *ArXiv: 1407.5107*, 2014.
- [28] P. Billingsley, *Convergence of probability measures*, ser. Wiley series in Probability and Statistics. New York: Wiley, 2009, vol. 493.
- [29] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1974.
- [30] A. Kuznetsov, "Solvable markov processes," Ph.D. dissertation, University of Toronto, 2004.