OPEN ACCESS

University of BRISTOL

Peer reviewed version

University of Bristol - Explore Bristol Research

General rights

DETECTING TRENDS IN TWITTER TIME SERIES

Tijl De Bie[1,2]*, Jefrey Lijffijt*[1]   *Cédric Mesnage*[2]*, Raúl Santos-Rodríguez*[2]

[1]Data Science Lab   [2]Dept of Engineering Mathematics
Ghent University – iMinds, Belgium   University of Bristol, United Kingdom

ABSTRACT

Detecting underlying trends in time series is important in many settings, such as market analysis (stocks, social media coverage) and system monitoring (production facilities, networks). Although many properties of the trends are common across different domains, others are domain-specific. In particular, modelling human activities such as their behaviour on social media, often leads to sharply defined events separated by periods without events. This paper is motivated by time series representing the number of tweets per day addressed to a specific Twitter user. Such time series are characterized by the combination of (1) an underlying trend, (2) concentrated bursts of activity that can be arbitrarily large, often attributable to an event, e.g., a tweet that goes viral or a real-world event, and (3) random fluctuations/noise. We present a new probabilistic model that accurately models such time series in terms of peaks on top of a piece-wise exponential trend. Fitting this model can be done by solving an efficient convex optimization problem. As an empirical validation of the approach, we illustrate how this model performs on a set of Twitter time series, each one addressing a particular music artist, which we manually annotated with events as a reference.

Index Terms— Trend detection, time series, convexity

1. INTRODUCTION

Trend detection from numerical time series is important in many settings: it is used for technical analysis of financial time series analysis (e.g. the stock market); for the detection of anomalous behaviour, e.g., in monitoring industrial plants such as offshore windfarms and chemical reactors, in which case the *deviation* from the trend is of interest; and for the analysis of time series that quantify the public's interest in a certain matter, e.g. derived from microblogs such as tweets. In the latter case, both the trend itself is relevant to understand longer-term shifts in the public's interest, as well as deviations from the trend, to detect particular events that could be of interest in their own right and that may be underlying these deviations. Despite the similar goals, each domain carries different underlying assumptions.

Twitter time series This paper is motivated by time series derived from Twitter, and more specifically the number of tweets per day about a certain 'topic'. This topic can be defined by a specific search query to the Twitter API. For example, in our experiments, each query is defined by the Twitter handle of a popular artist listed on the website `reverbnation.com`, and we retrieve all tweets that directly address this artist. The Twitter time series then consists of the number of such tweets per day. An example is shown in Figure 1.
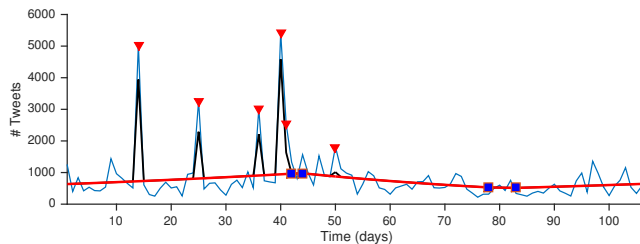

Fig. 1. Number of tweets per day (blue line), trend (red line), trend plus peaks (black line, occluded by red when there are no peaks). Peak positions are indicated by red triangles, and changes in trend indicated by blue squares.

Twitter time series contain rich information about the public interest enjoyed by a topic: the trend reveals the baseline amount of public interest the topic attracts, while the peaks deviating from the trend are often symptomatic of real-world or viral virtual events that boost the public's interest for a very short time period, often less than 1 day. Importantly, peaks are (almost) always *positive*: it is hard to imagine a situation that would cause the number of tweets about a certain topic to be dramatically lower on a specific day.

Contributions The main contribution in this paper is an efficient algorithm to separate trend from peaks in a Twitter time series. Although we use Twitter time series as motivating example, we expect that the proposed trend detection method is applicable to any time series with the following properties:

- The measurements are counts.
- The majority of measurements are close to a slowly

varying trend, potentially affected by periodic fluctuations (e.g., seasonal or weekday effects).

- Very large *positive* deviations from the trend (*peaks*) are possible, but sparse (i.e., occurring at few time points).

- Negative peaks are not possible.

These assumptions are formalised in Section 2, leading to a principled probabilistic interpretation for the trend, peaks, and observed time series.

The L_1 filter [1] is arguably the state-of-the-art method in trend detection. This and other related works are discussed in Section 3. In Section 4, we demonstrate the potential of the proposed method, illustrating that (1) it compares favourably with the L_1 filter, (2) it effectively extracts visual trends in a subjective evaluation on Twitter time series, and (3) events detected by the method correspond to events that were identified by detailed inspection of the Twitter data.

2. A SIMPLE GRAPHICAL MODEL FOR TWITTER TIME SERIES

2.1. Measurements, trend, and peaks

Let us denote the time series by means of a vector $\mathbf{y} \in \mathbb{N}^T$ with $T \in \mathbb{N}$ the number of time points, i.e., assuming all measurements to be positive integers. We will refer to the measurement at time t as y_t. For simplicity, we shall assume that time points are equidistant, though our results are easy to generalise should this not be the case.

We postulate the existence of a trend, denoted using a vector $\mathbf{x} \in \mathbb{R}_+^T$, and its t'th value as x_t. Additionally, we postulate a peak time series denoted $\mathbf{z} \in \mathbb{R}_+^T$, with z_t denoting the t'th value in this series.

While the meaning of $\mathbf{y}$ is obvious - it contains the observed number of tweets on the consecutive time points - let us clarify the meaning of $\mathbf{x}$ and $\mathbf{z}$. First note that y_t, being a count, can be modeled well as a Poisson random variable, assuming it is possible to estimate the Poisson rate parameter. We take x_t to represent that Poisson rate parameter *in the absence of a peak*. Since the expected value of a Poisson random variable is equal to its rate parameter, this means that the expected value of y_t is equal to x_t whenever there is no peak at time t, so $y_t \sim \text{Poisson}(x_t)$, where $\text{Poisson}(\lambda)$ signifies a Poisson distribution with rate parameter λ, i.e. for which $P(k) = \frac{\lambda^k}{k!} \exp(-\lambda)$.

We argue that the effect of a peak is best modelled as a multiplicative effect. To be precise, the same peak value z_t should have the same multiplicative effect on the expected value of y_t, regardless of the trend x_t. (In the example of music artists, this would mean that the release of a comparable music video should multiply their expected number of tweets by the same factor.) This can be modeled by assuming that the rate parameter in the presence of a peak is simply the product
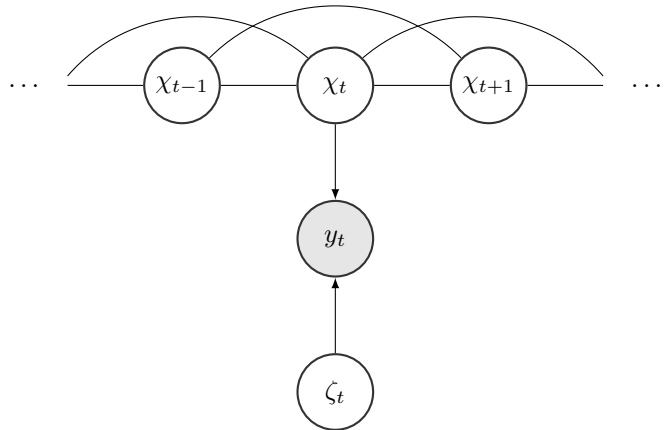

Fig. 2. Probabilistic graphical model for a Twitter time series. Shaded nodes represent observed variables while non-shaded nodes represent hidden variables.

of x_t and z_t, in other words, so $y_t \sim \text{Poisson}(x_t z_t)$. Thus, the absence of a peak at time t, can be modelled by ensuring that $z_t = 1$.

For reasons that will become clear below, it will prove convenient to use the logarithms of the trend values $\mathbf{x}$, denoted as $\boldsymbol{\chi}$ with $\chi_t \triangleq \log(x_t)$, as well as the logarithms of the peaks $\mathbf{z}$, denoted as $\boldsymbol{\zeta}$ with $\zeta_t \triangleq \log(z_t)$. Then, the above equation can be rewritten in additive terms as:

$$y_t \sim \text{Poisson}(\exp(\chi_t + \zeta_t)). \tag{1}$$

This is the essence of our probabilistic model for the time series. It can be represented as the graphical model in Fig. 2. The graph is partially directed (also called *chain graph*), with a mix of directed and undirected edges. In terms of factor graphs, there is a factor for each clique in the undirected part (prior on $\boldsymbol{\chi}$), another factor for the prior of ζ_t, and finally, a factor for each y_t conditioned on χ_t and ζ_t. In the next subsection, we will detail how the remaining assumptions spelled out in Sec. 1 can be formalized in terms of priors on $\boldsymbol{\zeta}$ and $\boldsymbol{\chi}$.

2.2. The priors on the trend and the peaks

Smoothness of the trend A natural way to model smoothly varying trends is by assuming a trend $\mathbf{x}$ that changes exponentially, i.e., its rate of growth (be it positive or negative) is constant. In the log domain, this means that $\boldsymbol{\chi}$ varies linearly over time. Of course, trends will occasionally change, and it is of particular interest to find out when that happens. One could argue however that trends whose growth rate vary frequently are less likely. A prior reflecting that would be $p_{\boldsymbol{\chi}}(\boldsymbol{\chi}) \propto \exp(-\lambda_1 ||\nabla^2 \boldsymbol{\chi}||_0)$, where $\lambda_1 \in \mathbb{R}^+$. Here, the linear operator ∇^2 represents the numerical approximation of the second order derivative, computed as $\chi_{t+1} - 2 \cdot \chi_t + \chi_{t-1}$ at time point t. Thus, $||\nabla^2 \boldsymbol{\chi}||_0$ represents the number of times the slope of $\boldsymbol{\chi}$ changes over the time range.

The use of the L_0 norm in the prior will ensure that log-trends χ with few slope changes are preferred, such that χ is piece-wise linear and thus $\mathbf{x}$ piece-wise exponential. L_0 norms are hard to handle computationally though. Fortunately, relaxing the L_0 norm to an L_1 norm leads to a much easier to handle prior, while still preferring sparsity of its argument $\nabla^2\chi$, such that the piece-wise exponential nature of $\mathbf{x}$ is maintained. Thus, as a prior for χ we will use:

$$
\begin{aligned}
p_{\boldsymbol{\chi}}(\boldsymbol{\chi}) &\propto \exp(-\lambda_1||\nabla^2\boldsymbol{\chi}||_1) \\
&= \prod_{t=2}^{T-1} \exp(-\lambda_1|\chi_{t-1} - 2\chi_t + \chi_{t+1}|).
\end{aligned}
$$

We note in passing that $p(\boldsymbol{\chi})$ is an improper prior and therefore its integral is unbounded. Even so, as we observe $\mathbf{y}$ and will be interested in estimating the trend $\mathbf{x}$ (as well as peaks), the use of an improper prior is justified and will cause no problems.

Sparse but possibly very large positive peaks We assume the presence of *positive peaks only*. This means that $\zeta_t \geq 0$, or equivalently that $z_t \geq 1$ for all t.

Additionally, we assume that the ζ_t (or equivalently, z_t) at different time points are *i.i.d.* This means that $p_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) = \prod_{t=1}^{T} p_{\zeta}(\zeta_t)$.

Regarding the *size of the peaks*: it is well known (e.g. Ch. 18 in [2]) that the amount of attention received by viral memes can be dramatic, and can therefore be modelled best by means of a Pareto (power law) distribution. We follow this approach, and model the peak values z_t as a Pareto distribution, $p_z(z_t) = \lambda_2 z_t^{-(\lambda_2+1)}$, with λ_2 the shape parameter.[1] Pareto distributions are heavy-tailed distributions.[2] This means that a large multiplicative effect of a peak on the observed number of tweets can still be attributed a large probability by our model. It will prove convenient to express this density function in the log domain, in terms of ζ_t. Considering the relation $z_t = \exp(\zeta_t)$, the density function in terms of this transformed variable $\zeta_t \geq 0$ is found as:

$$
\begin{aligned}
p_{\zeta}(\zeta_t)d\zeta_t &= p_z(z_t)dz_t, \\
\Rightarrow p_{\zeta}(\zeta_t) &= p_z(\exp(\zeta_t)) \cdot \frac{d\exp(\zeta_t)}{d\zeta_t}, \\
&= \lambda_2 \exp(-(\lambda_2+1)\zeta_t)\exp(\zeta_t), \\
&= \lambda_2 \exp(-\lambda_2\zeta_t).
\end{aligned}
$$

In other words, a prior of a Pareto distribution on z_t is equivalent to an exponential prior on ζ_t with rate parameter λ_2.

In Sec. 2.4 we will prove that this particular prior also induces *sparsity*: many peak values ζ_t will be equal to 0 in the maximum likelihood estimate for χ and ζ. More specifically,

we will establish a relation between the shape parameter λ_2 and the sparsity of $\boldsymbol{\zeta}$.

2.3. Maximum likelihood estimation of the trend and peaks

We are now ready to formulate the joint probability model:

$$
\begin{aligned}
p(\mathbf{y},\boldsymbol{\chi},\boldsymbol{\zeta}) &= p_{\boldsymbol{\chi}}(\boldsymbol{\chi}) \prod_{t=1}^{T} \left(p_{\chi}(\zeta_t)p_{y|\chi,\zeta}(y_t|\zeta_t,\chi_t) \right), \\
&\propto \exp\left(-\lambda_1||\nabla^2\boldsymbol{\chi}||_1\right) \prod_{t=1}^{T} \exp\left(-\lambda_2\zeta_t\right) \\
&\quad \cdot \frac{\exp\left((\chi_t + \zeta_t)y_t\right)}{y_t!} \exp\left(-\exp(\chi_t + \zeta_t)\right).
\end{aligned}
$$

Taking minus the logarithm yields the negative log-likelihood, up to an additive constant:

$$
-\log(p(\mathbf{y},\boldsymbol{\chi},\boldsymbol{\zeta})) = \lambda_1||\nabla^2\boldsymbol{\chi}||_1 -
$$
$$
\sum_{t=1}^{T} \left(-\lambda_2\zeta_t + (\chi_t + \zeta_t)y_t - \log(y_t!) - \exp(\chi_t + \zeta_t)\right).
$$

Thus, ignoring terms that are constant w.r.t. χ and ζ, the maximum likelihood estimation of the trend χ and peaks ζ amounts to solving the following optimization problem:

$$
\begin{aligned}
\min_{\boldsymbol{\chi},\boldsymbol{\zeta}} \quad & \lambda_1||\nabla^2\boldsymbol{\chi}||_1 + \\
& \sum_{t=1}^{T} \left(\lambda_2\zeta_t - (\chi_t + \zeta_t)y_t + \exp(\chi_t + \zeta_t)\right) \\
\text{s.t.} \quad & \boldsymbol{\zeta} \geq 0.
\end{aligned} \tag{2}
$$

This is a convex optimization problem, which can be solved e.g. using interior point solvers such as those used by CVX.[3]

2.4. Properties

We list below four important properties of the proposed trend detection method, which provide more insight and facilitate the tuning of the parameters λ_1 and λ_2.

Property 2.1 (The trend $\mathbf{x}$ is piece-wise exponential) *It is well known that when minimising an L_1 norm the solution usually enforces sparsity on its argument (i.e., many zero elements) [3]. In our case, this implies that the second order derivative of the maximum likelihood trend $\nabla^2\chi$ will have many zero elements, which means that the estimated χ is piecewise linear. Therefore, this amounts to a piece-wise exponential trend $\mathbf{x}$.*

We mentioned earlier that our aim is to identify possibly large but sparse peaks. Our model naturally deals with large peaks due to its use of a Pareto distribution for z_t. However, sparsity does not follow directly. The following theorem provides the desired result:

---

[1] The scale parameter is equal to 1.

[2] Given a random variable $A \sim$ Exponential(λ), a random variable $B \sim$ Pareto(α), and an arbitrarily large factor $r \in \mathbb{R}$, there always exists a value $s \in \mathbb{R}$ such that $P(B > s) > r \cdot P(A > s)$.

[3] http://cvxr.com/cvx/

Theorem 2.2 *Let ζ be the minimizer of the optimization problem (2). For any t for which $y_t \leq \lambda_2$, it holds that $\zeta_t = 0$.*

Proof We make use of the KKT conditions [4], which are both necessary and sufficient for this optimization problem. Let us introduce a KKT multiplier ν_t for the constraint $\zeta_t \geq 0$. Then the KKT conditions are given as:

Stationarity w.r.t. ζ_t : $\lambda_2 - y_t + \exp(\chi_t + \zeta_t) - \nu_t = 0$,

Primal feasibility : $\zeta_t \geq 0$,

Dual feasibility : $\nu_t \geq 0$,

Comp. slackness : $\zeta_t \nu_t = 0$.

For $y_t \leq \lambda_2$, it holds that $\lambda_2 - y_t + \exp(\chi_t + \zeta_t) > 0$, such that the stationarity condition can only be satisfied if $\nu_t > 0$. Thus, in order to satisfy the complementary slackness condition, $\zeta_t = 0$. ∎

In other words, for any t for which $y_t \leq \lambda_2$, it holds that for the maximum likelihood solution $\zeta_t = 0$. Note that the converse does not necessarily hold. Even so, it allows for an easy way to set λ_2 to a sensible value.

We now state and prove two theorems regarding the limit cases for the smoothness parameter λ_1.

Theorem 2.3 *Let ζ and χ be the joint minimizers of the optimization problem (2), and let $\lambda_2 > 0$. Then for $\lambda_1 \to 0$, $\exp(\chi_t) \to y_t$ and $\zeta_t \to 0$.*

Proof This follows from the KKT stationarity condition and complementary slackness conditions for ζ_t (see proof of Theorem 2.2), along with the stationarity condition for χ_t:

$$y_t = \exp(\chi_t + \zeta_t).$$

Plugging this expression for y_t into the stationarity condition for ζ_t yields:
$$\lambda_2 - \nu_t = 0,$$

such that $\nu_t = \lambda_2 > 0$ and thus from the complementary slackness condition it must follow that $\zeta_t = 0$. Thus, $y_t = \exp(\chi_t)$. ∎

This means that for λ_1 sufficiently small, no peaks will be detected (regardless of λ_2) and, given that $x_t \triangleq \exp(\chi_t)$ (by definition), the expected value of the Poisson random variables y_t will become exactly equal to their observations.

Theorem 2.4 *Let ζ and χ be the joint minimizers of the optimization problem (2). Then, for $\lambda_1 \to \infty$, $||\nabla^2 \chi||_1 \to 0$.*

Proof Assume to the contrary that $||\nabla^2 \chi||_1 > \delta > 0$ for λ_1 arbitrarily large. Then the objective function in problem (2) becomes infinitely large for $\lambda_1 \to \infty$. On the other hand, choosing χ such that $||\nabla^2 \chi||_1 = 0$ leads to a finite minimum. Indeed, it can be upper bounded by T by setting $\zeta = \chi = 0$. Thus, for the minimizer χ, it holds that $||\nabla^2 \chi||_1 \to 0$ for $\lambda_1 \to 0$. ∎

In other words, by choosing λ_1 sufficiently large, it is possible to ensure that the resulting trend $\mathbf{x}$ is a single exponential (or a linear function in the log-domain).

2.5. Incorporating periodic trends

The approach can be extended to consider time series that involve seasonal or otherwise periodic trends. In our particular application, it made sense to consider weekday variations as a confounding factor, and hence we incorporated that into the model. Given a period of k, this is done by including an additional variable vector $\boldsymbol{\pi} \in \mathbb{R}^k$, and modelling the count y_t as a Poisson random variable with rate parameter $\exp(\chi_t + \zeta_t + \pi_{\mod(t-1,k)+1})$, with the added constraint that $\sum_{i=1}^{k} \pi_i = 0$ to ensure identifiability. The complexity of the optimization problem is hardly affected by this. If desired, a prior on $\boldsymbol{\pi}$ could be added as well, which could make sense if k is not much smaller than T.

3. RELATED WORK

Many trend detection methods have been proposed in the literature. For a detailed review on different trend extraction techniques we refer the reader to [5].

Hodrick-Prescott filter Also known as L_2 filter [6], this approach assumes an additive model $y_t = x_t + n_t$, where n_t is a random component. The L_2 filter consists in determining the trend $\mathbf{x}$ by minimizing the following objective function,

$$\lambda ||\nabla^2 \mathbf{x}||_2^2 + ||\mathbf{y} - \mathbf{x}||_2^2$$

where $\lambda > 0$ is the regularization parameter which controls the trade-off between the smoothness of $\mathbf{x}$ and the residual $(y_t - x_t)$. From a probabilistic perspective, this is equivalent to assuming that n_t has a Gaussian prior distribution.

L_1 **filtering and variants** By replacing $||\nabla^2 \mathbf{x}||_2^2$ with $||\nabla^2 \mathbf{x}||_1$ (i.e. replacing the square of the L_2 norm with the L_1 norm), we obtain the popular L_1 filter [1]. Although no probabilistic interpretation was offered by Kim et al. [1], it is easy to check that it amounts to maximum likelihood estimation when the residuals are modelled as Gaussian. Importantly, they also mention in passing a possible extension to include the detection of peaks. This is done by introducing a peak vector $\mathbf{z}$ similar to ours, and adding an extra term to the optimization problem to ensure it is small in L_1 norm, such that the objective function becomes:

$$\lambda_1 ||\nabla^2 \mathbf{x}||_1 + \lambda_2 ||\mathbf{z}||_1 + ||\mathbf{y} - (\mathbf{x} + \mathbf{z})||_2^2.$$

Here the term $\lambda_1 ||\nabla^2 \mathbf{x}||_1$ ensures a piece-wise linear trend, $\lambda_2 ||\mathbf{z}||_1$ ensures sparsity of the peaks, and $||\mathbf{y} - (\mathbf{x} + \mathbf{z})||_2^2$ ensures a small squared error between the trend+peaks and

the measured time-series. It is possible to attach a probabilistic interpretation to this time-series as follows: an improper prior on the trend, ensuring a piece-wise linear trend; an exponential distribution for the absolute value of the peaks; Gaussian noise between trend+peaks and the measured signal. The authors do not study this objective in any great detail, and, although this formulation bears some similarities with our objective function, it would be suboptimal in our setting because (1) a piece-wise linear trend makes little sense for a time-series that is always positive, (2) it seems reasonable to assume that the probability of a peak should depend on its height *relative to the trend*, rather than on its height in absolute terms, (3) Gaussian noise is only a good approximation for the Poisson distribution when the measured counts are sufficiently large, and (4) their method considers both positive and negative peaks. The first two issues can be addressed by using their method on the log-transformed time series, and exponentiating the resulting trend and trend+peak values. However, doing that aggravates the third problem: the logarithm of a Poisson random variable cannot be modelled well using a Gaussian distribution. This can easily be seen by observing that a Poisson random variable may equal 0 with a high probability (and in any case with a non-zero probability), while the probability attributed to $-\infty = \log(0)$ will always be zero under any Gaussian approximation. The last issue can be fixed by adding the constraint $z_t >= 0$.

4. Experiments

To demonstrate the potential of the method we will first study how it compares to its closest competitor method on synthetic data. Then, we will show how it effectively extracts visual trends while the detected peaks correspond to events that were identified by detailed inspection of the Twitter data.

4.1. Synthetic time series

In order to be as fair as possible, we implemented a modified version of the standard L_1 filter so that it allows for (only positive) peaks and works with the log-transformed time series. Although then both methods behave similarly when the counts are large enough, we want to discuss their performance when this is not the case (i.e. when gaussian noise is no longer a good approximation for the Poisson distribution).

Data generation We generate 10 time series with 100 time points, following a Poisson distribution as described in (1). In order to simplify the analysis, we consider just one trend with no changes of slope, $\chi_t = log(15) - 0.01t$. We keep the values of χ_t small to simulate reasonably small counts. The time series contain three peaks randomly positioned within the first half of the time series. The height of the three peaks is the same and takes values in $\{0, 1, 2, 3\}$ (also in the log-domain and therefore determining ζ_t).

Table 1. Average number of false positives for each method.

Method	Peak Height	λ_2					
		0	3	6	9	12	15
Probabilistic	0	97.0	16.9	3.8	0.5	0.0	0.0
	1	97.0	18.6	3.4	0.2	0.0	0.0
	2	97.0	14.0	2.3	0.4	0.2	0.0
	3	97.0	15.7	2.9	0.6	0.1	0.0
L_1 *filtering*	0	97.0	31.7	6.7	0.9	0.2	0.0
	1	97.0	32.4	6.9	1.1	0.1	0.0
	2	97.0	25.6	4.3	0.4	0.0	0.0
	3	97.0	29.1	6.0	0.7	0.0	0.0

Table 2. Average number of false negatives for each method.

Method	Peak Height	λ_2					
		0	3	6	9	12	15
Probabilistic	0	0.0	2.4	2.9	3.0	3.0	3.0
	1	0.0	0.0	0.0	0.1	0.2	0.9
	2	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0	0.0	0.0
L_1 *filtering*	0	0.0	2.0	3.0	3.0	3.0	3.0
	1	0.0	0.0	0.0	0.2	1.7	2.8
	2	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0	0.0	0.0

Results From now on, our method is referred to as *probabilistic*. We set λ_1 just high enough for both methods to guarantee that the result contains just one trend. We then vary λ_2 to quantify its influence. We analyse the detection of peaks in terms of the average number of false positives (finding a peak when there is none) in Table 1, and false negatives (not finding a peak when there is one) in Table 2, respectively. The proposed method performs better than the modified L_1 filter, both in terms of false positives and false negatives.

4.2. Twitter time series

We applied our method to study trends using Twitter time series corresponding to several music artists.

Data collection We gathered the Twitter handles from 19,630 UK-based artists listed on reverbnation.com for whom a Twitter handle is specified. We used Twitter's API to search for all tweets about these Twitter handles, collecting 6,959,940 tweets from December 3rd 2014 to March 26th 2015, tweeted by 1,359,660 individuals. Once the dataset was generated, we manually annotated the time series, identifying peaks and looking for sensible explanations for the corresponding events. This was done independently of the fitting of the trend and the detection of the peaks, so that we could use the annotations as ground truth to validate the method. Due to the lack of space, we show a small subset of these time series.

Parameter tuning In order to choose a pair of parameter values that suited all the Twitter time series, we used both
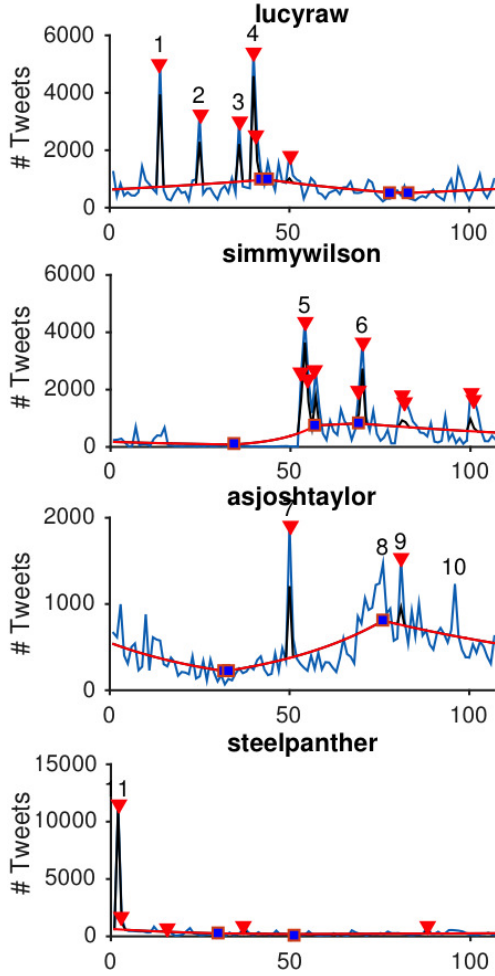
Fig. 3. Number of tweets per day (blue line), trend (red line), trend plus peaks (black line, occluded by red when there are no peaks) for 4 artists. Peak positions are indicated by red triangles, and changes in trend indicated by blue squares.

Th. 2.2 and our prior knowledge to set λ_2 to the 80th percentile of **y**, so there can be at most 20% of peaks.

Results Fig. 3 shows the different Twitter time series (number of tweets per day, in blue) corresponding to 4 artists in our dataset. Numbers are attached to the peaks that were manually annotated by the authors. After running the method, we checked the peaks found by both the human annotator and the automatic method to give some insight into the meaning of the these peaks. Peaks 1 to 4 correspond to *lucyraw* actively interacting on Twitter. Two more peaks were detected, the first being closely related to peak 4, while the last one did not relate to any actual event (i.e., false positive). In peak 5, *simmywilson* released a picture that became very popular, while in peak 6 she sparked her fan base's interest tweeting about ageing. The remaining detected peaks did not match any further events. *Asjoshtaylor*'s collaboration with *Desmond Mary*

caused peak 7 and peak 9 happened to be his birthday. Interestingly, we could not find events associated with peaks 8 and 10, but the human annotator interpreted those as peaks. On the other hand, the automatic method was able to correctly avoid those false negatives. In peak 11 *Steelpanther* uploaded a photograph with a very popular band (*5SOS*), profiting from their notoriety. Although the human annotator did not find any more peaks, the automatic method found events that corresponding to a TV interview and a gig. In general, the proposed method compares positively with the human annotator. The automatic method detected most of the peaks that were annotated by hand, but shows false positives. Additionally, it also correctly identifies events that the human annotator overlooked. However, the main advantage of the automatic method is its scalability: the human annotator is able to deal with tagging a few time series hand but would struggle when faced with thousands or millions.

5. CONCLUSION

Trend detection methods are efficient tools to infer the tendency in time series. However, a good estimator of the trend depends on the problem at hand. We show that we can define a principled probabilistic framework encompassing our beliefs regarding the model to forecast the trend of bursty Twitter time series in a very simple way and we apply this to extract the trends corresponding to different music artists.

6. References

[1] S-J Kim, K. Koh, S. Boyd, and D. Gorinevsky, "*l*1 trend filtering," *SIAM review*, vol. 51, no. 2, pp. 339–360, 2009.

[2] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

[3] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC, 2015.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[5] T. Alexandrov, S. Bianconcini, E.B. Dagum, P. Maass, and T.S. McElroy, "A review of some modern approaches to the problem of trend extraction," *Econometric Reviews*, vol. 31, no. 6, pp. 593–624, 2012.

[6] R.J. Hodrick and E.C. Prescott, "Postwar U.S. business cycles: An empirical investigation," *Journal of Money, Credit and Banking*, vol. 29, no. 1, pp. 1–16, 1997.