University of Vermont

ScholarWorks @ UVM

College of Engineering and Mathematical Sciences Faculty Publications

College of Engineering and Mathematical Sciences

12-1-2020

The shocklet transform: a decomposition method for the identification of local, mechanism-driven dynamics in sociotechnical time series

David Rushing Dewhurst
University of Vermont

Thayer Alshaabi
University of Vermont

Dilan Kiley
Zillow Group

Michael V. Arnold
University of Vermont

Joshua R. Minot
University of Vermont

See next page for additional authors

Follow this and additional works at: https://scholarworks.uvm.edu/cemsfac

Part of the Human Ecology Commons, and the Medicine and Health Commons

Recommended Citation

Dewhurst DR, Alshaabi T, Kiley D, Arnold MV, Minot JR, Danforth CM, Dodds PS. The shocklet transform: a decomposition method for the identification of local, mechanism-driven dynamics in sociotechnical time series. EPJ Data Science. 2020 Dec 1;9(1):3.

This Article is brought to you for free and open access by the College of Engineering and Mathematical Sciences at ScholarWorks @ UVM. It has been accepted for inclusion in College of Engineering and Mathematical Sciences Faculty Publications by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.

Authors

David Rushing Dewhurst, Thayer Alshaabi, Dilan Kiley, Michael V. Arnold, Joshua R. Minot, Christopher M. Danforth, and Peter Sheridan Dodds

This article is available at ScholarWorks @ UVM: https://scholarworks.uvm.edu/cemsfac/14

Dewhurst et al. *EPJ Data Science* (2020) 9:3
https://doi.org/10.1140/epjds/s13688-020-0220-x

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

The shocklet transform: a decomposition method for the identification of local, mechanism-driven dynamics in sociotechnical time series

David Rushing Dewhurst[1,2*], Thayer Alshaabi[1,3], Dilan Kiley[4], Michael V. Arnold[1,2], Joshua R. Minot[1,2], Christopher M. Danforth[1,2] and Peter Sheridan Dodds[1,2]

*Correspondence:
david.dewhurst@uvm.edu
[1]Computational Story Lab, Vermont Complex Systems Center, University of Vermont, Burlington, United States
[2]Department of Mathematics & Statistics, University of Vermont, Burlington, United States
Full list of author information is available at the end of the article

Abstract

We introduce a qualitative, shape-based, timescale-independent time-domain transform used to extract local dynamics from sociotechnical time series—termed the Discrete Shocklet Transform (DST)—and an associated similarity search routine, the Shocklet Transform And Ranking (STAR) algorithm, that indicates time windows during which panels of time series display qualitatively-similar anomalous behavior. After distinguishing our algorithms from other methods used in anomaly detection and time series similarity search, such as the matrix profile, seasonal-hybrid ESD, and discrete wavelet transform-based procedures, we demonstrate the DST's ability to identify mechanism-driven dynamics at a wide range of timescales and its relative insensitivity to functional parameterization. As an application, we analyze a sociotechnical data source (usage frequencies for a subset of words on Twitter) and highlight our algorithms' utility by using them to extract both a typology of mechanistic local dynamics and a data-driven narrative of socially-important events as perceived by English-language Twitter.

Keywords: Nonparametric statistics; Sociotechnical time series; Time-domain filtering; Social media

0 Introduction

The tasks of peak detection, similarity search, and anomaly detection in time series is often accomplished using the discrete wavelet transform (DWT) [1] or matrix-based methods [2, 3]. For example, wavelet-based methods have been used for outlier detection in financial time series [4], similarity search and compression of various correlated time series [5], signal detection in meteorological data [6], and homogeneity of variance testing in time series with long memory [7]. Wavelet transforms have far superior localization in the time domain than do pure frequency-space methods such as the short-time Fourier transform [8]. Similarly, the chirplet transform is used in the analysis of phenomena displaying periodicity-in-perspective (linearly- or quadratically-varying frequency), such as images

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Springer

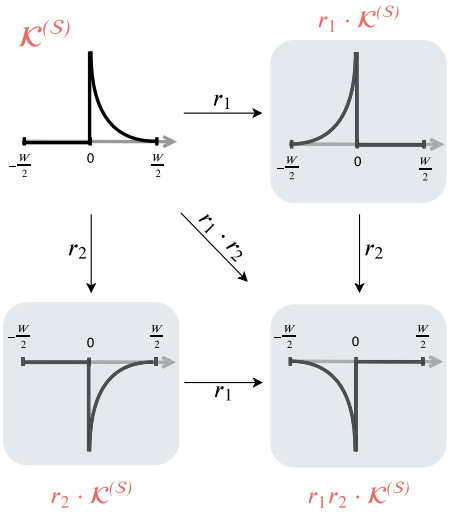

Figure 1 The discrete shocklet transform is generated through cross-correlation of pieces of shocks. this figure displays effects of the action of group elements $r_i \in R_4$ on a base "shock-like" kernel $\mathcal{K}$. The kernel $\mathcal{K}$ captures the dynamics of a constant lower level of intensity before an abrupt increase to a relatively high intensity which decays over a duration of $W/2$ units of time. By applying elements of R_4, we can effect a time reversal (r_1) and abrupt cessation of intensity followed by asymptotic convergence to the prior level of intensity (r_2), as well as the combination of these effects ($r_3 = r_1 \cdot r_2$)

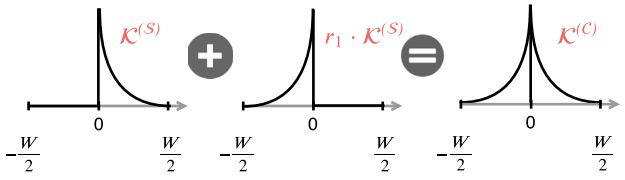

Figure 2 Shock dynamics. This figure provides a schematic for the construction of more complicated shock dynamics from a simple initial shape ($\mathcal{K}^{(S)}$). By acting on a kernel with elements r_i of the reflection group R_4 and function concatenation, we create shock-like dynamics, as exemplified by the symmetric shocklet kernel $\mathcal{K}^{(C)} = \mathcal{K}^{(S)} \oplus [r_1 \cdot \mathcal{K}^{(S)}]$ in this figure. In Sect. 2.3 we illuminate a typology of shock dynamics derived from combinations of these basic shapes

and radar signals [9–12]. Thus, when analyzing time series that are partially composed of exogenous shocks and endogenous shock-like local dynamics, we should use a small sample of such a function—a "shock", examples of which are depicted in Fig. 1, and functions generated by concatenation of these building blocks, such as that shown in Fig. 2.

In this work, we introduce the Discrete Shocklet Transform (DST), generated by cross-correlation functions of a shocklet. As an immediate example (and before any definitions or technical discussion), we contrast the DWT with the DST of a sociotechnical time series—popularity of the word "trump" on the social media website Twitter—in Fig. 3, which is a visual display of what we claim is the DST's suitability for detection of local mechanism-driven dynamics in time series.

We will show that the DST can be used to extract shock and shock-like dynamics of particular interest from time series through construction of an indicator function that compresses time-scale-dependent information into a single spatial dimension using prior information on timescale and parameter importance. Using this indicator, we are able to highlight windows in which underlying mechanistic dynamics are hypothesized to contribute a stronger component of the signal than purely stochastic dynamics, and demonstrate an algorithm—the Shocklet Transform and Ranking (STAR) algorithm—by which we are able to automate *post facto* detection of endogenous, mechanism-driven dynamics. As a complement to techniques of changepoint analysis, methods by which one can detect

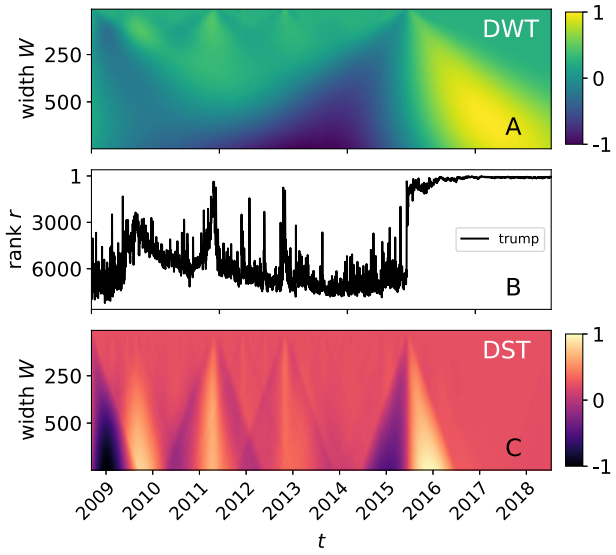

Figure 3 A comparison between the standard discrete wavelet transform (DWT) and our discrete shocklet transform (DST) of a sociotechnical time series. Panel (**B**) displays the daily time series of the rank r_t of the word "trump" on Twitter. As a comparison with the DST, we computed the DWT of r_t using the Ricker wavelet and display it in panel (**A**). Panel (**C**) shows the DST of the time series using a symmetric power shock, $\mathcal{K}^{(S)}(\tau|W,\theta) \sim \mathbf{rect}(\tau)\tau^{\theta}$, with exponent $\theta = 3$. We chose to compare the DST with the DWT because the DWT is similar in mathematical construction (see Appendix 1 for a more extensive discussion of this assertion), but differs in the choice of convolution kernel (a wavelet, in the case of the DWT, and a piece of a shock, in the case of the DST) and the method by which the transform accounts for signal at multiple timescales

changes in the level of time series [13, 14], the DST and STAR algorithm detect changes in the underlying mechanistic local dynamics of the time series. Finally, we demonstrate a potential usage of the shocklet transform by applying it to the LabMT Twitter dataset [15] to extract word usage time-series matching the qualitative form of a shock-like kernel at multiple timescales.

1 Data and theory

1.1 Data

Twitter is a popular micro-blogging service that allows users to share thoughts and news with a global community via short messages (up to 140 or, from around November 2017 on, 280 characters, in length). We purchased access to Twitter's "decahose" streaming API and used it to collect a random 10% sample of all public tweets authored between September 9, 2008 and April 4, 2018 [16]. We then parsed these tweets to count appearances of words included in the LabMT dataset, a set of roughly 10,000 of the most commonly used words in English [15]. The dataset has been used to construct nonparametric sentiment analysis models [17] and forecast mental illness [18] among other applications [19–21]. From these counts, we analyze the time series of word popularity as measured by rank of word usage: on day t, the most-used word is assigned rank 1, the second-most assigned rank 2, and so on to create time series of word rank r_t for each word.

1.2 Theory

1.2.1 Algorithmic details: description of the method

There are multiple fundamentally-deterministic mechanistic models for local dynamics of sociotechnical time series. Nonstationary local dynamics are generally well-described by exponential, bi-exponential, or power-law decay functions; mechanistic models thus usually generate one of these few functional forms. For example, Wu and Huberman described a stretched-exponential model for collective human attention [22], and Candia *et al.* derived a biexponential function for collective human memory on longer timescales [23]. Crane and Sornette assembled a Hawkes process for video views that produces power-law behavior by using power-law excitement kernels [24], and Lorenz-Spreen *et al.* demonstrated a speeding-up dynamic in collective social attention mechanisms [25], while De Domenico and Altmann put forward a stochastic model incorporating social heterogeneity and influence [26], and Ierly and Kostinsky introduced a rank-based, signal-extraction method with applications to meteorology data [27]. In Sect. 1.2.2 we conduct a literature review, contrasting our methods with existing anomaly detection and similarity search time series data mining algorithms and demonstrating that the DST and associated STAR algorithm differ substantially from these existing algorithms. We have open-sourced implementations of the DST and STAR algorithm; code for these implementations is available at a publicly-accessible repository.[a]

We do not assume any specific model in our work. Instead, by default we define a kernel $\mathcal{K}^{(\cdot)}$ as one of a few basic functional forms: exponential growth,

$$\mathcal{K}^{(S)}(\tau|W,\theta) \sim \text{rect}(\tau - \tau_0)e^{\theta(\tau-\tau_0)}; \tag{1}$$

monomial growth,

$$\mathcal{K}^{(S)}(\tau|W,\theta) \sim \text{rect}(\tau - \tau_0)\tau^{\theta}; \tag{2}$$

power-law decay,

$$\mathcal{K}^{(S)}(\tau|W,\theta) \sim \text{rect}(\tau - \tau_0)|\tau - \tau_0 + \varepsilon|^{-\theta}, \tag{3}$$

or sudden level change (corresponding with a changepoint detection problem),

$$\mathcal{K}^{(Sp)}(\tau|W,\theta) \sim \text{rect}(\tau - \tau_0)\big[\Theta(\tau) - \Theta(-\tau)\big], \tag{4}$$

where $\Theta(\cdot)$ is the Heaviside step function. The function rect is the rectangular function ($\text{rect}(x) = 1$ for $0 < x < W/2$ and $\text{rect}(x) = 0$ otherwise), while in the case of the power-law kernel we add a constant ε to ensure nonsingularity. The parameter W controls the support of $\mathcal{K}^{(\cdot)}(\tau|W,\theta)$; the kernel is identically zero outside of the interval $[\tau - W/2, \tau + W/2]$. We define the window parameter W as follows: moving from a window size of W to a window size of $W + \Delta W$ is equivalent to upsampling the kernel signal by the factor $W + \Delta W$, applying an ideal lowpass filter, and downsampling by the factor W. In other words, if the kernel function $\mathcal{K}^{(\cdot)}$ is defined for each of W linearly spaced points between $-N/2$ and $N/2$, moving to a window size of W to $W + \Delta W$ is equivalent to computing $\mathcal{K}^{(\cdot)}$ for each of $W + \Delta W$ linearly-spaced points between $-N/2$ and $N/2$. This holds the

dynamic range of the kernel constant while accounting for the dynamics described by the kernel at all timescales of interest. We enforce the condition that $\sum_{t=-\infty}^{\infty} \mathcal{K}^{(\cdot)}(t|W,\theta) = 0$ for any window size W.

It is decidedly not our intent to delve into the question of how and why deterministic underlying dynamics in sociotechnical systems arise. However, we will provide a brief justification for the functional forms of the kernels presented in the last paragraph as scaling solutions to a variety of parsimonious models of local deterministic dynamics:

- If the time series $x(t)$ exhibits exponential growth with a state-dependent growth damper $D(x)$, the dynamics can be described by

$$\frac{dx(t)}{dt} = \frac{\lambda}{D(x(t))}x(t), \qquad x(0) = x_0. \tag{5}$$

  If $D(x) = x^{1/n}$, the solution to this IVP scales as $x(t) \sim t^n$, which is the functional form given in Eq. (2). When $D(x) \propto 1$ (i.e., there is no damper on growth) then the solution is an exponential function, the functional form of Eq. (1).
- If instead the underlying dynamics correspond to exponential decay with a time- and state-dependent half-life $\mathcal{T}$, we can model the dynamics by the system

$$\frac{dx(t)}{dt} = -\frac{x(t)}{\mathcal{T}(t)}, \qquad x(0) = x_0, \tag{6}$$

$$\frac{d\mathcal{T}(t)}{dt} = f\big(\mathcal{T}(t), x(t)\big), \qquad \mathcal{T}(0) = \mathcal{T}_0. \tag{7}$$

  If f is particularly simple and given by $f(\mathcal{T},x) = c$ with $c > 0$, then the solution to Eq. (6) scales as $x(t) \sim t^{-1/c}$, the functional form of Eq. (3). The limit $c \to 0^+$ is singular and results in dynamics of exponential decay, given by reversing time in Eq. (1) (about which we expound later in this section).
- As another example, the dynamics could be essentially static except when a latent variable φ changes state or moves past a threshold of some sort:

$$\frac{dx(t)}{dt} = \delta\big(\varphi(t) - \varphi^*\big), \qquad x(0) = x_0, \tag{8}$$

$$\frac{d\varphi(t)}{dt} = g\big(\varphi(t), x(t)\big), \qquad \varphi(0) = \varphi_0. \tag{9}$$

  In this case the dynamics are given by a step function from x_0 to $x_0 + 1$ the first time $\varphi(t)$ changes position relative to φ^*, and so on; these are the dynamics we present in Eq. (4).

This list is obviously not exhaustive and we do not intend it to be so.

We can use kernel functions $\mathcal{K}^{(\cdot)}$ as basic building blocks of richer local mechanistic dynamics through function concatenation and the operation of the two-dimensional reflection group R_4. Elements of this group correspond to $r_0 = \mathrm{id}$, $r_1 =$ reflection across the vertical axis (time reversal), $r_2 =$ negation (e.g., from an increase in usage frequency to a decrease in usage frequency), and $r_3 = r_1 \cdot r_2 = r_2 \cdot r_1$. We can also model new dynamics by concatenating kernels, i.e., "glueing" kernels back-to-back. For example, we can generate "cusplets" with both anticipatory and relaxation dynamics by concatenating a shocklet

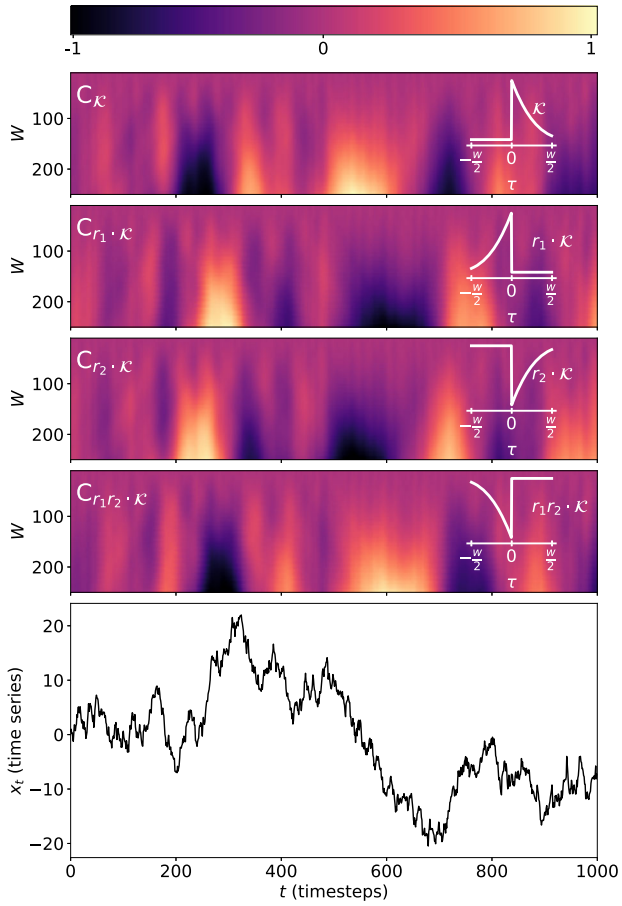

Figure 4 Effects of the reflection group R_4 on the shocklet transform. The top four panels display the results of the shocklet transform of a random walk $x_t = x_{t-1} + z_t$ with $z_t \sim \mathcal{N}(0,1)$, displayed in the bottom panel, using the kernels $r_j \cdot \mathcal{K}^{(S)}$, where $r_j \in R_4$

$\mathcal{K}^{(S)}$ with a time-reversed copy of itself:

$$\mathcal{K}^{(C)}(\tau|W,\theta) \sim \mathcal{K}^{(S)}(\tau|W,\theta) \oplus \left[r_1 \cdot \mathcal{K}^{(S)}(\tau|W,\theta)\right]. \tag{10}$$

We display an example of this concatenation operation in Fig. 2. For much of the remainder of the work, we conduct analysis using this symmetric kernel.

The discrete shocklet transform (DST) of the time series $x(t)$ is defined by

$$C_{\mathcal{K}^{(S)}}(t, W|\theta) = \sum_{\tau=-\infty}^{\infty} x(\tau + t)\mathcal{K}^{(S)}(\tau|W,\theta), \tag{11}$$

which is the cross-correlation of the sequence and the kernel. This defines a $T \times N_W$ matrix containing an entry for each point in time t and window width W considered.

To convey a visual sense of what the DST looks like when using a shock-like, asymmetric kernel, we compute the DST of a random walk $x_t - x_{t-1} = z_t$ (we define $z_t \sim \mathcal{N}(0,1)$) using a kernel function $\mathcal{K}^{(S)}(\tau|W,\theta) \sim \text{rect}(\tau)\tau^\theta$ with $\theta = 3$ and display the resulting matrix for window sizes $W \in [10, 250]$ in Fig. 4.

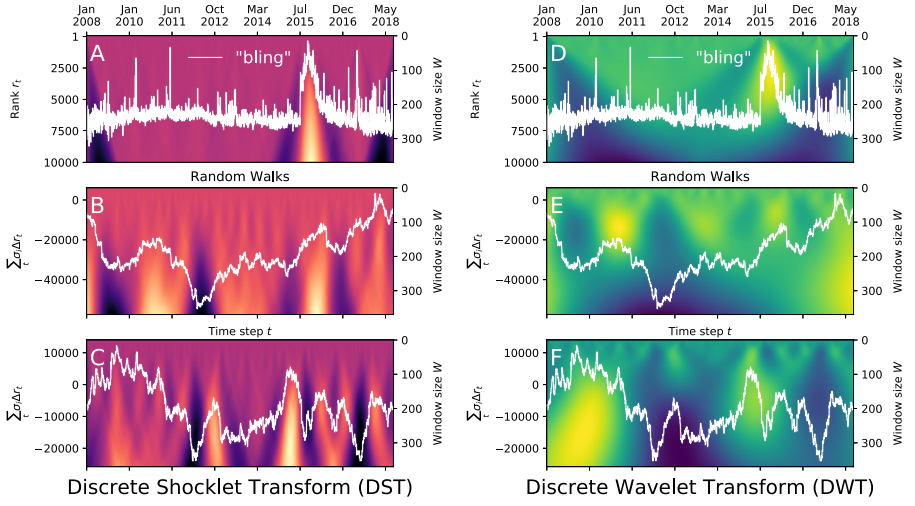

Figure 5 Intricate dynamics of sociotechnical time series. Panels (**A**) and (**D**) show the time series of the ranks down from top of the word "bling" on Twitter. Until mid-summer 2015, the time series presents as random fluctuation about a steady, relatively-constant level. However, the series then displays a large fluctuation, increases rapidly, and then decays slowly after a sharp peak. The underlying mechanism for these dynamics was the release of a popular song titled "Hotline Bling". To demonstrate the qualitative difference of the "bling" time series from draws from a null random walk model, the details of which are given in Appendix 1. Panels (**A**), (**B**), and (**C**) show the discrete shocklet transform of the original series for "bling" and the random walks $\sum_{t'\leq t}\Delta r_{\sigma_j t}$, showing the responsiveness of the DST to nonstationary local dynamics and its insensitivity to dynamic range. Panels (**D**), (**E**), and (**F**), on the other hand, display the discrete wavelet transform of the original series and of the random walks, demonstrating the DWT's comparatively less-sensitive nature to local shock-like dynamics

The effects of time reversal by action of r_1 are visible when comparing the first and third panels with the second and fourth panels, and the result of negating the kernel by acting on it with r_2 is apparent in the negation of the matrix values when comparing the first and second panels and with the third and fourth. For this figure, we used a random walk as an example time series here as there is, by definition, no underlying generative mechanism causing any shock-like dynamics; these dynamics appear only as a result of integrated noise. We are equally likely to see large upward-pointing shocks as large downward-pointing shocks because of this, which allows us to see the activation of both upward-pointing and downward-pointing kernel functions.

As a comparison with this null example, we computed the DST of a sociotechnical time series, the rank of the word "bling" among the LabMT words on Twitter, and two draws from a null random walk model, and displayed the results in Fig. 5. Here, we calculated the DST using the symmetric kernel given in Eq. (10). (For more statistical details of the null model, see Appendix 1.) We also computed the DWT of each of these time series and display the resulting wavelet transform matrices next to the shocklet transform matrices in Fig. 5. Direct comparison of the sociotechnical time series (r_t) with the draws from the null models reveals r_t's moderate autocovariance as well as the large, shock-like fluctuation that occurs in late July of 2015. (This underlying driver of this fluctuation was the release of a popular song entitled "Hotline Bling" on July 31st, 2015.) In comparison, the draws from the null model have a covariance with much more prominent time scaling and do not exhibit dramatic shock-like fluctuations as does r_t. Comparing the DWT of these time

series with the respective DST provides more evidence that the DST exhibits superior space-time localization of shock-like dynamics than does the DWT.

To aggregate deterministic behavior across all timescales of interest, we define the shock indicator function as the function

$$\mathrm{C}_{\mathcal{K}^{(S)}}(t|\theta) = \sum_{W} \mathrm{C}_{\mathcal{K}^{(S)}}(t, W|\theta)p(W|\theta), \tag{12}$$

for all windows W considered. The function $p(W|\theta)$ is a probability mass function that encodes prior beliefs about the importance of particular values of W. For example, if we are interested primarily in time series that display shock- or shock-like behavior that usually lasts for approximately one month, we might specify $p(W|\theta)$ to be sharply peaked about $W = 28$ days. Throughout this work we take an agnostic view on all possible window widths and so set $p(W|\theta) \propto 1$, reducing our analysis to a strictly maximum-likelihood based approach. Summing over all values of the shocklet parameter θ defines the shock indicator function,

$$\mathrm{C}_{\mathcal{K}^{(S)}}(t) = \sum_{\theta} \mathrm{C}_{\mathcal{K}^{(S)}}(t|\theta)p(\theta) \tag{13}$$

$$= \sum_{\theta,W} \mathrm{C}_{\mathcal{K}^{(S)}}(t, W|\theta)p(W|\theta)p(\theta). \tag{14}$$

In analogy with $p(W\theta)$, the function $p(\theta)$ is a probability density function describing our prior beliefs about the importance of various values of θ. As we will show later in this section, and graphically in Fig. 6, the shock indicator function is relatively insensitive to choices of θ possessing a nearly-identical ℓ_1 norm for wide ranges of θ and different functional forms of $\mathcal{K}^{(S)}$.

After calculation, we normalize $\mathrm{C}_{\mathcal{K}^{(S)}}(t)$ so that it again integrates to zero and has $\max_t \mathrm{C}_{\mathcal{K}^{(S)}}(t) - \min_t \mathrm{C}_{\mathcal{K}^{(S)}}(t) = 2$. The shock indicator function is used to find windows in which the time series displays anomalous shock- or shock-like behavior. These windows are defined as

$$\left\{t \in [0, T] : \text{intervals where } \mathrm{C}_{\mathcal{K}^{(S)}}(t) \geq s\right\}, \tag{15}$$

where the parameter $s > 0$ sets the sensitivity of the detection.

The DST is relatively insensitive to quantitative changes to its functional parameterization; it is a qualitative tool to highlight time periods of unusual events in a time series. In other words, it does not detect statistical anomalies but rather time periods during which the time series appears to take on certain qualitative characteristics without being too sensitive to a particular functional form. We analyzed two example sociotechnical time series—the rank of the word "bling" on Twitter (for reasons we will discuss presently)—and the price time series of Bitcoin (symbol BTC) [28], the most actively-used cryptocurrency [29], and of one null model, a pure random walk. For each time series, we computed the shock indicator function using two kernels, each of which had a different functional form (one kernel given by the function of Eq. (10) and one of the identical form but constructed by setting $\mathcal{K}^{(S)}(\tau|W, \theta)$ to the function given in Eq. (1)), and evaluating each kernel over a wide range of its parameter θ. We also vary the maximum window size from

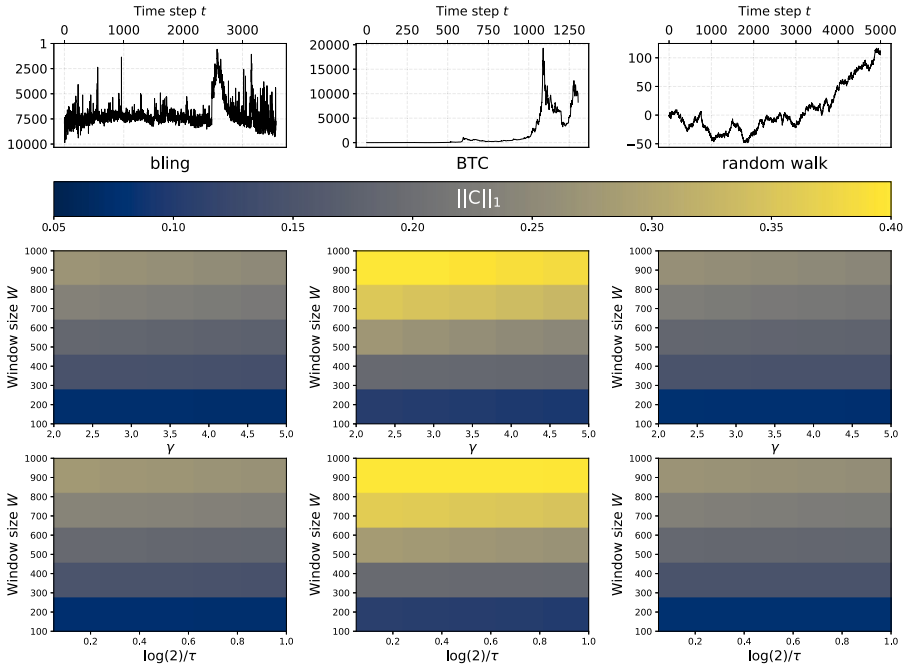

Figure 6 Parameters Sweep. The shock indicator function is relatively insensitive to functional forms $\mathcal{K}^{(\cdot)}$ and values of the kernel's parameter vector θ so long as the kernel functions are qualitatively similar (e.g., for cusp-like dynamics—as considered in this figure and in Eq. (10)—$\mathcal{K}^{(C)}$ displaying increasing rates of increase followed by decreasing rates of decrease). Here we have computed the shock indicator function $\mathbf{C}_{\mathcal{K}^{(S)}}(\tau|\theta)$ (Eq. (12)) for three different time series: two sociotechnical and one null example. From left to right, the top row of figures displays the rank usage time series of the word "bling" on Twitter, the price of the cryptocurrency Bitcoin, and a simple Gaussian random walk. Below each time series we display parameter sweeps over combinations of $(\theta, W_{\max})$ for two kernel functions: one kernel given by the function of Eq. (10) and another of the identical form but constructed by setting $\mathcal{K}^{(S)}(\tau|W,\theta)$ to the function given in Eq. (1). The ℓ_1 norms of the shock indicator function are nearly invariant across the values of the parameters θ for which we evaluated the kernels. However, the shock indicator function does display dependence on the maximum window size $W_{\max}$, with large $W_{\max}$ associated with larger ℓ_1 norm. This is because a larger window size allows the DST to detect shock-like behavior over longer periods of time

$W = 100$ to $W = 1000$ to explore the sensitivity of the shock indicator function to this parameter. We display the results of this comparative analysis in Fig. 6. For each time series, we plot the ℓ_1 norm of the shock indicator function for each (θ, W) combination. We find that, as stated earlier in this section, the shock indicator function is relatively insensitive to both functional parameterization and value of the parameter θ; for any fixed W, the ℓ_1 norm of the shock indicator function barely changed regardless of the value of θ or choice of $\mathcal{K}^{(\cdot)}$. However, the maximum window size does have a notable effect on the magnitude of the shock indicator function; higher values of W are associated with larger magnitudes. This is a reasonable finding, since higher maximum W means that the DST is able to capture shock-like behavior that occurs over longer timespans and hence may have values of higher magnitude over longer periods than for comparatively lower maximum W.

That the shock indicator function is a relative quantity is both beneficial and problematic. The utility of this feature is that the dynamic behavior of time series derived from systems of widely-varying time and length scales can be directly compared; while the rank of a word on Twitter and—for example—the volume of trades in an equity security are entirely different phenomena measured in different units, their shock indicator functions

are unitless and share similar properties. On the other hand, the Shock Indicator Function carries with it no notion of dynamic range. Two time series x_t and y_t could have identical shock indicator functions but have spans differing by many orders of magnitude, i.e., $\operatorname{diam} x_t \equiv \max_t x_t - \min_t x_t \gg \operatorname{diam} y_t$. (In other words, the diameter of a time series in interval I is just the dynamic range of the time series over that interval.) We can directly compare time series inclusive of their dynamic range by computing a weighted version of the shock indicator function, $C_{\mathcal{K}}(t)\Delta x(t)$, which we term the weighted shock indicator function (WSIF). A simple choice of weight is

$$\Delta x(t) = \operatorname*{diam}_{t\in[t_b,t_e]} x_t, \tag{16}$$

where t_b and t_e are the beginning and end times of a particular window. We use this definition for the remainder of our paper, but one could easily imagine using other weighting functions, e.g., maximum percent change (perhaps applicable for time series hypothesized to increment geometrically instead of arithmetically).

These final weighted shock indicator functions are the ultimate output of the shocklet transform and ranking (STAR) algorithm; the weighting corresponds to the actual magnitude of the dynamics and constitutes the "ranking" portion of the algorithm, while the weighting will only be substantially larger than zero if there existed intervals of time during which the time series exhibited shock-like behavior as indicated in Eq. (15). We present a conceptual, bird's-eye view of the STAR algorithm (of which the DST is a core component) in Fig. 7. Though this diagram is lacking in technical detail, we have included it in an effort to provide a bird's-eye view of the entire STAR algorithm and to help orient the reader on the conceptual process underpinning the algorithm.

1.2.2 Algorithmic details: *comparison with existing methods*

On a coarse scale, there are five nonexclusive categories of time series data mining tasks [30]: similarity search (also termed indexing), clustering, classification, summarization, and anomaly detection. The STAR algorithm is a qualitative, shape-based, timescale-independent, similarity search algorithm. As we have shown in the previous section, the discrete shocklet transform (a core part of the overarching STAR algorithm) is *qualitative*, meaning that it does not depend too strongly on values of functional parameters or even the functions used in the cross-correlation operation themselves, as long as the functions share the same qualitative dynamics (e.g., increasing rates of increase followed by decreasing rates of decrease for cusp-like dynamics); hence, it is primarily *shape-based* rather than relying on the quantitative definition of a particular functional form. STAR is *timescale-independent* as it is able to detect shock-like dynamics over a wide range of timescales limited only by the maximum window size for which it is computed. Finally, we believe that it is best to categorize STAR as a *similarity search* algorithm as this seems to be the best-fitting label for STAR that is given in the five categories listed at the beginning of this section; STAR is designed for searching within sociotechnical time series for dynamics that are similar to the shock kernel in some way, albeit similar in a qualitative sense and over any arbitrary timescale, not functionally similar in numerical value and characteristic timescale. However, it could also be considered a type of qualitative, shape-based anomaly detection algorithm because we are searching for behavior that is, in some sense, anomalous compared to a usual baseline behavior of many time series (though see discussion

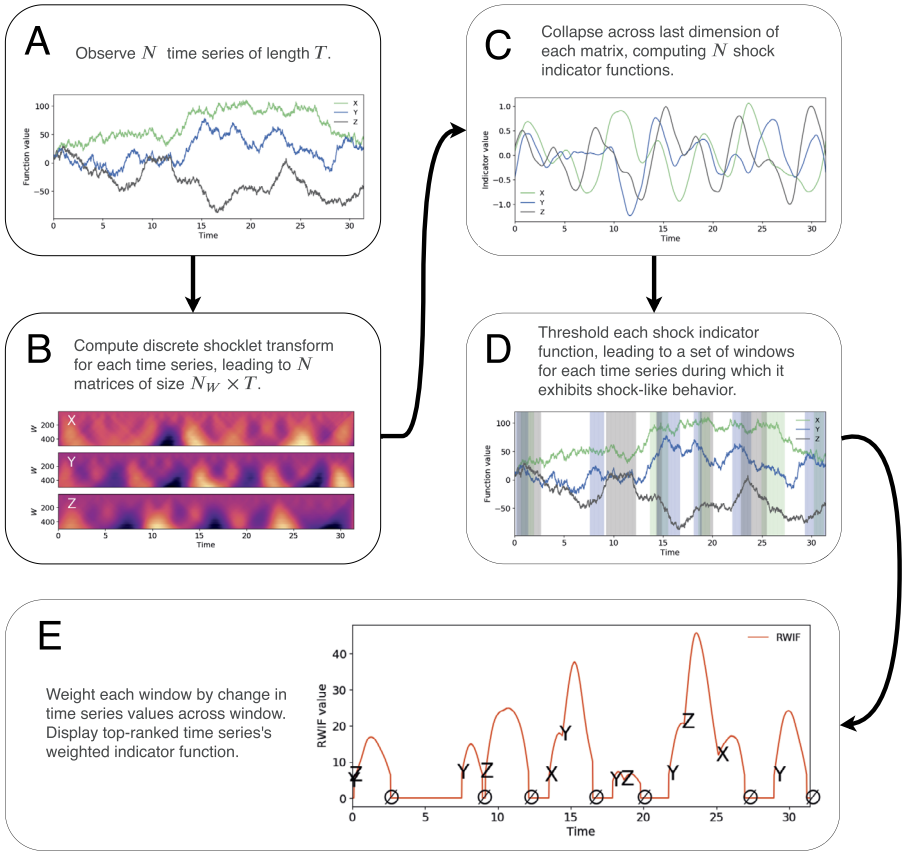

Figure 7 STAR. The Shocklet Transform And Ranking (STAR) algorithm combines the discrete shocklet transform (DST) with a series of transformations that yield intermediate results, such as the cusp indicator function (panel (**C**) in the figure) and windows during which each univariate time series displays shock-like behavior (panel (**D**) in the figure). Each of these intermediate results is useful in its own right, as we show in Sect. 2. We display the final output of the STAR algorithm, a univariate indicator that condenses information about which of the time series exhibits the strongest shock-like behavior at each point in time

at the beginning of the anomaly detection subsection near the end of this section: STAR is an algorithm that can detect defined anomalous behavior, *not* an algorithm to detect arbitrary statistical anomalies).

As such, we are unaware of any existing algorithm that satisfies these four criteria and believe that STAR represents an entirely new class of algorithms for sociotechnical time series analysis. Nonetheless, we now provide a detailed comparison of the DST with other algorithms that solve related problems, and in Sect. 2.1 provide an in-depth quantitative comparison with another nonparametric algorithm (Twitter's anomaly detection algorithm) that one could attempt to use to extract shock-like dynamics from sociotechnical time series.

Similarity search—here the objective is to find time series that minimize some similarity criterion between candidate time series and a given reference time series. Algorithms to solve this problem include nearest-neighbor methods (e.g., k-nearest neighbors [31] or a locality-sensitive hashing-based method [32, 33]), the discrete Fourier and wavelet transforms [5, 34–36]; and bit-, string-, and matrix-based representations [30, 37–39]. With suitable modification, these algorithms can also be used to solve time series clustering

problems. Generic dimensionality-reduction techniques, such as singular value decomposition/principal components analysis [40–42], can also be used for similarity search by searching through a dataset of lower dimension. Each of these classes of algorithms differs substantially in scope from the discrete shocklet transform. Chief among the differences is the focus on the entire time series. While the discrete shocklet transform implicitly searches the time series for similarity with the kernel function at all (user-defined) relevant timescales and returns qualitatively-matching behavior at the corresponding timescale, most of the algorithms considered above do no such thing; the user must break the time series into sliding windows of length τ and execute the algorithm on each sliding window; if the user desires timescale-independence, they must then vary τ over a desired range. An exception to this statement is Mueen's subsequence similarity search algorithm (MSS) [43], which computes sliding dot products (cross-correlations) between a long time series of length T and a shorter kernel of length M before defining a Euclidean distance objective for the similarity search task. When this sliding dot product is computed using the fast Fourier transform, the computational complexity of this task is $\mathcal{O}(T \log T)$. This computational step is also at the core of the discrete shocklet transform, but is performed for multiple kernel function arrays (more precisely, for the kernel function resampled at multiple user-defined timescales). Unlike the discrete shocklet transform, MSS does not subsequently compute an indicator function and does not have the self-normalizing property, while the matrix profile algorithm [39] computes an indicator function of sorts (their "matrix profile") but is not timescale-independent and is quantitative in nature; it does not search for a qualitative shape match as does the discrete shocklet transform. We are unaware of a similarity-search algorithm aside from STAR that is both qualitative in nature and timescale-independent.

Clustering—given a set of time series, the objective is to group them into groups, or clusters, that are more homogeneous within each cluster than between clusters. Viewing a collection of N time series of length T as a set of vectors in $\mathbb{R}^T$, any clustering method that can be effectively used on high-dimensional data has potential applicability to clustering time series. Some of these general clustering methods include k-means and k-medians algorithms [44–46], hierarchical methods [47–49], and density-based methods [47, 50–52]. There are also methods designed for clustering time series data specifically, such as error-in-measurement models [53], hidden Markov models [54], simulated annealing-based methods [55], and methods designed for time series that are well-fit by particular classes of parametric models [56–59]. Although the discrete shocklet transform component of the STAR algorithm could be coerced into performing a clustering task by using different kernel functions and elements of the reflection group, clustering is not the intended purpose of the discrete shocklet transform or STAR more generally. In addition, none of the clustering methods mentioned replicate the results of the STAR algorithm. These clustering methods uncover groups of time series that exhibit similar behavior over their entire domain; application of clustering methods to time series subsequences carries leads to meaningless results [60]. Clustering algorithms are also shape-independent in the sense that they cluster data into groups that share similar features, but do not search for specific known features or shapes in the data. In contrast with this, when using the STAR algorithm we already have specified a specific shape—for example, the shock shape demonstrated above—and are searching the data across timescales for occurrences of that shape. The STAR algorithm also does not require multiple time series in order to function

effectively, differing from any clustering algorithm in this respect; a clustering algorithm applied to $N = 1$ data points trivially returns a single cluster containing the single data point. The STAR algorithm operates identically on one or many time series as it treats each time series independently.

Classification—classification is the canonical supervised statistical learning problem in which data x_i is observed along with a discrete label y_i that is taken to be a function of the data, $y_i = f(x_i) + \varepsilon$; the goal is to recover an approximation to f that precisely and accurately reproduces the labels for new data [61]. This is the category of time series data mining algorithms that least corresponds with the STAR algorithm. The STAR algorithm is unsupervised—it does not require training examples ("correct labels") in order to find subsequences that qualitatively match the desired shape. As above, the STAR algorithm also does not require multiple time series to function well, while (non-Bayesian) classification algorithms rely on multiple data points in order to learn an approximation to f.[b]

Summarization—since time series can be arbitrarily large and composed of many intricately-related features, it may be desirable to have a summary of their behavior that encompasses the time series's "most interesting" features. These summaries can be numerical, graphical, or linguistic in nature. Underlying methodologies for time series summary tasks include wavelet-based approaches [62, 63], genetic algorithms [64, 65], fuzzy logic and systems [66–68], and statistical methods [69]. Though intermediate steps of the STAR algorithm can certainly be seen as a time series summarization mechanism (for example, the matrix computed by the DShT or the weighted shock indicator functions used in determinning rank relevance of individual time series at different points in time), the STAR algorithm was not designed for time series summarization and should not be used for this task as it will be outperformed by essentially any other algorithm that was actually designed for summarization. Any "summary" derived from the STAR algorithm will have utility only in summarizing segments of the time series the behavior of which match the kernel shape, or in distinguishing segments of the time series that do have a similar shape as the kernel from ones that do not.

Anomaly detection—if a "usual" model can be defined for the system under study, an anomaly detection algorithm is a method that finds deviations from this usual behavior. Before we briefly review time series anomaly detection algorithms and compare them with the STAR algorithm, we distinguish between two subtly different concepts: this data mining notion of anomaly detection, and the physical or social scientific notion of anomalous behavior. In the first sense, *any* deviation from the "ordinary" model is termed an anomaly and marked as such. The ordinary model may not be a parametric model to which the data is compared; for example, it may be implicitly defined as the behavior that the data exhibits most of the time [70]. In physical and social sciences, on the other hand, it may be observed that, given a particular set of laboratory or observational conditions, a material, state vector, or collection of agents exhibits phenomena that is anomalous when compared to a specific reference situation, even if this behavior is "ordinary" for the conditions under which the phenomena is observed. Examples of such anomalous behavior in physics and economics include: spectral behavior of polychromatic waves that is very unusual compared to the spectrum of monochromatic waves (even though it is typical for polychromatic waves near points where the wave's phase is singular) [71]; the entire concept of anomalous diffusion, in which diffusive processes with mean square displacement (autocovariance functions) scaling as $\langle r(t) \rangle \sim t^{\alpha}$ are said to diffuse anomalously if

$\alpha \not\approx 1$ (since $\alpha = 1$ is the scaling of the Wiener process's autocovariance function) [72, 73], even though anomalous diffusion is the rule rather than the exception in intra-cellular and climate dynamics, as well as financial market fluctuations; and behavior that deviates substantially from the "rational expectations" of non-cooperative game theory, even though such deviations are regularly observed among human game players [74, 75]. This distinction between algorithms designed for the task of anomaly detection and algorithms or statistical procedures that test for the existence of anomalous behavior, as defined here, is thus seen to be a subtle but significant difference. The DST and STAR algorithm fall into the latter category: the purpose for which we designed the STAR algorithm is to extract windows of anomalous behavior as defined by comparison with a particular null qualitative time series model (absence of clear shock-like behavior), not to perform the task of anomaly detection writ large by indicating the presence of arbitrary samples or dynamics in a time series that does not in some way comport with the statistics of the entire time series.

With these caveats stated, it is not the case that there is no overlap between anomaly detection algorithms and algorithms that search for some physically-defined anomalous behavior in time series; in fact, as we show in Sect. 2.1, there is some significant convergence between windows of shock-like behavior indicated by STAR and windows of anomalous behavior indicated by Twitter's anomaly detection algorithm when the underlying time series exhibits relatively low variance. Statistical anomaly detection algorithms typically propose a semi-parametric model or nonparametric test and confront data with the model or test; if certain datapoints are very unlikely under the model or exceed certain theoretical boundaries derived in constructing the test, then these datapoints are said to be anomalous. Examples of algorithms that operate in this way include: Twitter's anomaly detection algorithm (ADV), which relies on generalized seasonal ESD test [76, 77]; the EGADS algorithm, which relies on explicit time series models and outlier tests [78]; time-series model and graph methodologies [79, 80]; and probabilistic methods [81, 82]. Each of these methods is strictly focused on solving the first problem that we outlined at the beginning of this subsection: that of finding points in one or more time series during which it exhibits behavior that deviates substantially from the "usual" or assumed behavior for time series of a certain class. As we outlined, this goal differs substantially from the one for which we designed STAR: searching for segments of time series (that may vary widely in length) during which the time series exhibits behavior that is qualitatively similar to underlying deterministic dynamics (shock-like behavior) that we believe is anomalous when compared to non-sociotechnical time series.

2 Empirical results

2.1 Comparison with Twitter's anomaly detection algorithm

Through the literature review in Sect. 1.2 we have demonstrated that, to our knowledge, there exists no algorithm that solves the same problem for which STAR was designed—to provide a qualitative, shape-based, timescale-independent measure of similarity between multivariate time series and a hypothesized shape generated by mechanistic dynamics. However, there are existing algorithms designed for nonparametric anomaly detection that could be used to alert to the presence of shock-like behavior in sociotechnical time series, which is the application for which we originally designed STAR. One leading example of such an algorithm is Twitter's Anomaly Detection Vector (ADV) algorithm.[c] This

algorithm uses an underlying statistical test, seasonal-hybrid ESD, to test for the presence of outliers in periodic and nonstationary time series [76, 77]. We perform a quantitative and qualitative comparison between the STAR and ADV to compare their effectiveness at the task for which we designed STAR—determining qualitative similarity between shock-like shapes over a wide range of timescales—and to contrast the signals picked up by each algorithm, which, as we show, differ substantially. Before presenting results of this analysis, we note that this comparison is not entirely fair; though ADV is a state-of-the-art anomaly detection algorithm, it was not designed for the task for which we designed STAR, and so it is not exactly reasonable to assume that ADV would perform as well as STAR on this task. In an attempt to ameliorate this problem, we have chosen a quantitative benchmark for which our *a priori* beliefs did not favor the efficacy of either algorithm.

As both STAR and ADV are unsupervised algorithms, we compare their quantitative performance by assessing their utility in generating features for use in a supervised learning problem. Since the macro-economy is a canonical example of a sociotechnical system, we consider the problem of predicting the probability of a U.S. economic recession using only a minimal set of indicators from financial market data. Models for predicting economic recessions variously use only real economic indicators [83–85], only financial market indicators [86, 87], or a combination of real and financial economic indicators [88, 89]. We take an approach that is both simple and relatively granular, focusing on the ability of statistics of individual equity securities to jointly model U.S. economic recession probability. For each of the equities that was in the Dow Jones Industrial Average between 1999-07-01 to 2017-12-31 (a total of $K = 32$ securities), we computed both the DST (outputting the shock indicator function), STAR algorithm (outputting windows of shock-like behavior), and the ADV routine on that equity's volume traded time series (number of shares transacted), which we sampled at a daily resolution for a total of $T = 6759$ observations for each security. We then fit linear models of the form

$$E\left[\log \frac{\mathbf{p}}{1-\mathbf{p}}\right] = \mathbf{X}\boldsymbol{\beta}, \tag{17}$$

where p_t is the recession probability on day t as given by the U.S. Federal Reserve (hence $\mathbf{p}$ is the length-T vector of recession probabilities).[d] When we the model represented by Eq. (17) using ADV or STAR as the algorithms generating features, the design matrix $\mathbf{X}$ is a binary matrix of shape $T \times (K+1)$ with entry X_{tk} equal to one if the algorithm indicated an anomaly or shock-like behavior respectively in security k at time t and equal to zero if it did not (the +1 in the dimensionality of the matrix corresponds to the prepended column of ones that is necessary to fit an intercept in the regression). When we fit the model using the shock indicator function generated by the DST, the matrix $\mathbf{X}$ is instead given by the matrix with column k equal to the shock indicator function of security k.

We evaluate the goodness of fit of these linear models using the proportion of variance explained (R^2) statistic; these results are summarized graphically in Fig. 8. The linear using ADV-indicated anomalies as features had $R^2_{\text{ADV}} = 0.341$, while the model using the shock indicator function as columns of the design matrix had $R^2_{\text{DST}} = 0.455$ and the model using STAR-indicated shocks as features had $R^2_{\text{STAR}} = 0.496$. This relative ranking of feature importance remained constant when we used model log-likelihood ℓ as the performance metric instead of R^2, with ADV, DST, and STAR respectively exhibiting $\ell_{\text{ADV}} = -16{,}278$,

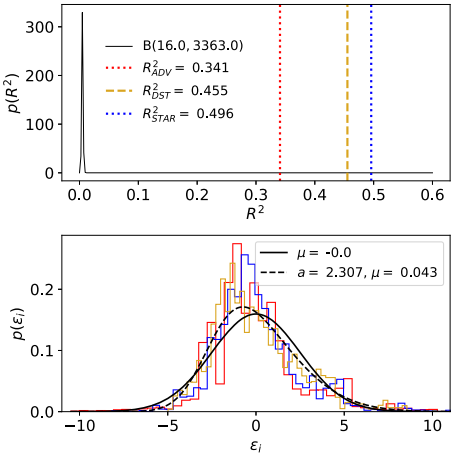

Figure 8 Analytical comparison of U.S. economic recession. We modeled the log odds ratio of a U.S. economic recession using three ordinary least squares regression models. Each model used one of the ADV method's anomaly indicator, the shock indicator function resulting from the discrete shocklet transform, and the windows of shock-like behavior output by the STAR algorithm as elements of the design matrix. The models that used features constructed by the DST or STAR outperformed the model that used features constructed by ADV as measured by both R^2 (displayed in the top panel) and model log-likelihood. The black curve in the top panel displays the null distribution of R^2 under the assumption that no regressor (column of the design matrix) actually belongs to the true linear model of the data [91, 92]. The lower panel displays the empirical probability distributions of the model residuals ε_i

$\ell_{\text{DST}} = -15{,}633$, and $\ell_{\text{STAR}} = -15{,}372$. Each linear model exhibited a distribution of residuals ε_t that did not drastically violate the zero-mean and distributional-shape assumptions of least-squares regression; a maximum likelihood fit of a normal probability density to the empirical error probability distribution $p(\varepsilon_t)$ gave mean and variance as $\mu = 0$ to within numerical precision and $\sigma^2 \approx 6.248$, while a maximum likelihood fit of a skew-normal probability density [90] to the empirical error probability distribution gave mean, variance, and skew as $\mu \approx 0.043$, $\sigma^2 \approx 6.025$, and $a \approx 2.307$. Taken in the aggregate, these results constitute evidence to suggest that features generated by the DST and STAR algorithms are superior in the task of classifying time periods as belonging to recessions or not than are features derived from the ADV method.

As a further comparison of the STAR algorithm and ADV, we generated anomaly windows (in the case of ADV) and windows of shock-like behavior (in the case of STAR) for the usage rank time series of each of the 10,222 words in the LabMT dataset. We computed the Jaccard similarity index for each word w (also known as the intersection over union) between the set of STAR windows $\{I_i^{\text{STAR}}(w)\}_i$ and the set of ADV windows $\{I_i^{\text{ADV}}(w)\}_i$,

$$J_w(\text{STAR}, \text{ADV}) = \frac{(\bigcup_i I_i^{\text{STAR}}(w)) \cap (\bigcup_i I_i^{\text{ADV}}(w))}{\bigcup_{j\in\{\text{STAR,ADV}\}} \bigcup_i I_i^j(w)}. \tag{18}$$

We display the word time series and ADV and STAR windows for a selection of words pertaining to the 2016 U.S. presidential election in Fig. 9. (These words display shock-like behavior in a time interval surrounding the election, as we demonstrate in the next section, hence our selection of them as examples here.)

We display the distribution of all Jaccard similarity coefficients in Fig. 10. Most words have relatively little overlap between anomaly windows returned by ADV and windows of shock-like dynamics returned by STAR, but there are notable exceptions. In particular, a review of the figures contained in the online index suggests that ADV's and STAR's windows overlap most when the shock-like dynamics are particularly strong and surrounded by a time series with relatively low variance; they agree the most when hypothesized underlying deterministic mechanics are strongest and the effects of noise are lowest. The pronounced spikes in the words "crooked" and "stein" in Fig. 9 are an example of this

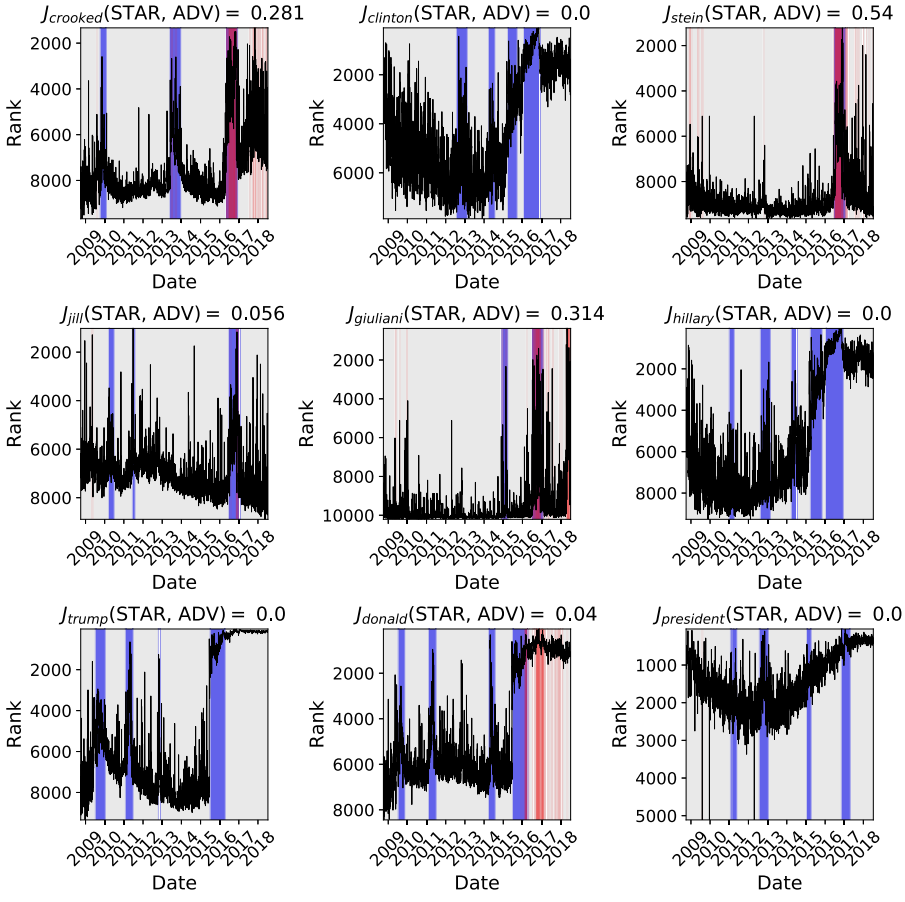

Figure 9 Comparison of STAR and Twitter's Anomaly Detection Vector (ADV) algorithm used for detecting phenomena in Twitter 1gram time series. The Jaccard similarity coefficient is presented for each 1-gram and the region where events on detected are shaded for the respective algorithm. Blue-shaded windows correspond with STAR windows of shock-like behavior, while red-shaded windows correspond with ADV windows of anomalous behavior (and hence purple windows correspond to overlap between the two). In general, ADV is most effective at detecting brief spikes or strong shock-like signals, whereas STAR is more sensitive to longer-term shocks and shocks that occur in the presence of surrounding noisy or nonstationary dynamic. ADV does not treat strong periodic fluctuations as anomalous by design; though this may or may not be a desirable feature of a similarity search or anomaly detection algorithm, it is certainly not a flaw in ADV but simply another differentiator between ADV and STAR

phenomenon. However, when the time series has high variance or exhibits strong nonstationarity, ADV often does not indicate that there are windows of anomalous behavior while STAR does indicate the presence of shock-like dynamics; the panels of the words "trump", "jill", and "hillary" in Fig. 9 demonstrate these behaviors.

Taken in the aggregate, these results suggest that a state-of-the-art anomaly detection algorithm, such as Twitter's ADV, and a qualitative, shape-based, timescale-independent similarity search algorithm, such as STAR, do have some overlapping properties but are largely mutually-complementary approaches to identifying and analyzing the behavior of sociotechnical time series. While ADV and STAR both identify strongly shock-like dynamics that occur when the surrounding time series has relatively low variance, their behavior diverges when the time series is strongly nonstationary or has high variance. In this case, ADV is an excellent tool for indicating the presence of strong outliers in the

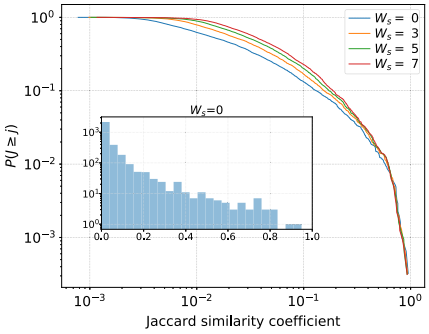

Figure 10 Jaccard similarity coefficients. Complimentary cumulative distribution function (CCDF) of Jaccard similarity coefficients for regions that Twitter's ADV and our STAR algorithm detect patterns or anomalies (see Fig. 9). Window sizes are varied to include $W_s \in \{0, 3, 5, 7\}$ (i.e. detections within $t_j \pm W_s$ are as part of the intersection). Time series with $J_{\mathbf{word}_j} = 0$ are omitted from the CCDF. The inset histogram shows the distribution of Jaccard similarity coefficients for $W_s = 0$ (i.e. exact matches), $J = 0$ time series are included

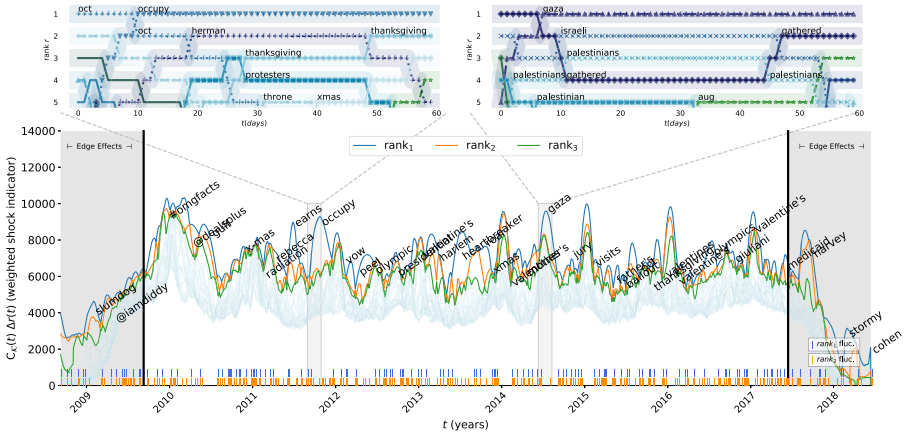

Figure 11 Shock leaderboard. Time series of the ranked and weighted shock indicator function. At each time step t, the weighted spike indicator functions (WSIF) are sorted so that the word with the highest WSIF corresponds to the top time series, the words with the second-highest WSIF corresponds to the second time series, and so on. Vertical ticks along the bottom mark fluctuations in the word occupying ranks 1 and 2 of WSIF values. Top panels present the ranks of WSIF values for words in the top 5 WSIF values in a given time step for the sub-sampled period of 60 days. An interactive version of this graphic is available at the authors' webpage: http://compstorylab.org/shocklets/ranked_shock_weighted_interactive.html

data, while STAR continues to indicate the presence of shock-like dynamics in a manner that is less sensitive to the time series's stationarity or variance.

2.2 Social narrative extraction

We seek both an understanding of the intertemporal semantic meaning imparted by windows of shock-like behavior indicated by the STAR algorithm and a characterization of the dynamics of the shocks themselves. We first compute the shock indicator and weighted shock indicator functions (WSIFs) for each of the 10,222 labMT words filtered from the gardenhose dataset, described in Sect. 1.1, using a power kernel with $\theta = 3$. At each point in time, words are sorted by the value of their WSIF. The jth highest valued WSIF at each temporal slice, when concatenated across time, defines a new time series. We perform this computation for the top ranked $k = 20$ words for the entire time under study. We also perform this process using the "spike" kernel of Eq. (4) and display each resulting time series in Fig. 11 (shock kernel) and Fig. 12 (spike kernel). (We term the spike kernel as such because we have $\frac{d\mathcal{K}^{(Sp)}(\tau)}{d\tau} = \delta(\tau)$ on the domain $[-W/2, W/2]$, the Dirac delta function; its

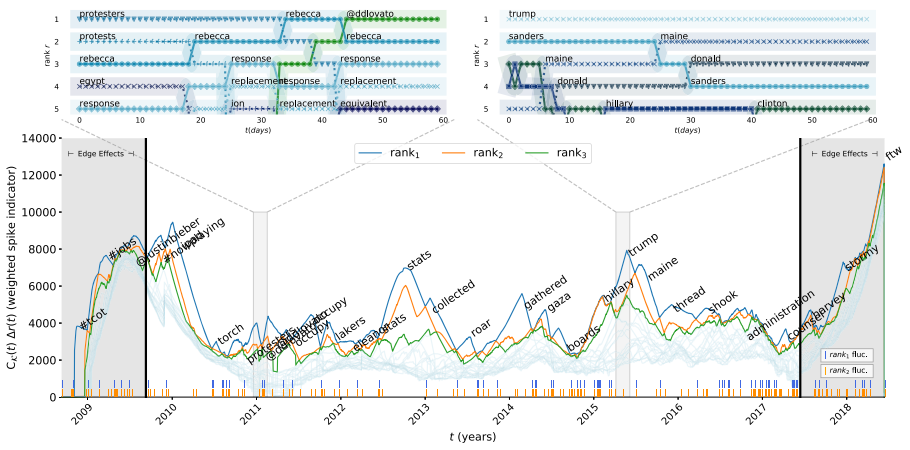

Figure 12 Spike leaderboard. Time series of the ranked and weighted spike indicator function. At each time step t, the weighted spike indicator functions (WSpIF) are sorted so that the word with the highest WSpIF corresponds to the top time series, the words with the second-highest WSpIF corresponds to the second time series, and so on. Vertical ticks along the bottom mark fluctuations in the word occupying ranks 1 and 2 of WSpIF values. Top panels present the ranks of WSpIF values for words in the top 5 WSpIF values in a given time step for the sub-sampled period of 60 days. The top left panel, demonstrates the competition for social attention between geopolitical concerns—street protests in Egypt—and popular artists and popular culture influence—Rebecca Black and Demi Lovato. The top right panel displays the language surrounding the 2016 U.S. presidential election immediately after Donald Trump announced his candidacy. An interactive version of this graphic is available at the authors' webpage:
http://compstorylab.org/shocklets/ranked_spike_weighted_interactive.html

underlying mechanistic dynamics are completely static except for one point in time during which the system is driven by an ideal impulse function.)

The $j = 1$ word time series is annotated with the corresponding word at relative maxima of order 40. (A relative maximum x_s of order k in a time series is a point that satisfies $x_s > x_t$ for all t such that $|t-s| \leq k$.) This annotation reveals a dynamic social narrative concerning popular events, social movements, and geopolitical fluctuation over the past near-decade. Interactive versions of these visualizations are available on the authors' website.[e] To further illuminate the often-turbulent dynamics of the top j ranked weighted shock indicator functions, we focus on two particular 60-day windows of interest, denoted by shading in the main panels of Figs. 11 and 12. In Fig. 11, we outline a period in late 2011 during which multiple events competed for collective attention:

- the 2012 U.S. presidential election (the word "herman", referring to Herman Cain, a presidential election contender);
- Occupy Wall Street protests ("occupy" and "protestors");
- and the U.S. holiday of Thanksgiving ("thanksgiving")

Each of these competing narratives is reflected in the top-left inset. In the top right inset, we focus on a time period during which the most distinct anomalous dynamics corresponded to the 2014 Gaza conflict with Israel ("gaza", "israeli", "palestinian", "palestinians", "gathered"). In Fig. 12, we also outline two periods of time: one, in the top left panel, demonstrates the competition for social attention between geopolitical concerns:

- street protests in Egypt ("protests", "protesters", "egypt", "response");
- and popular artists and popular culture ("rebecca", referring to Rebecca Black, a musician, and "@ddlovato", referring to another musician, Demi Lovato).

In the top right panel we demonstrate that the most prominent dynamics during late 2015 are those of the language surrounding the 2016 U.S. presidential election immediately after Donald Trump announced his candidacy ("trump", "sanders", "donald", "hillary", "clinton", "maine").

We note that these social narratives uncovered by the STAR algorithm might not emerge if we used a different algorithm in an attempt to extract shock-like dynamics in sociotechnical time series. We have already shown (in the previous section) that at least one state-of-the-art anomaly detection algorithm is unlikely to detect abrupt, shock-like dynamics that occur in time series that are nonstationary or have high variance. We display side-by-side comparisons of the indicator windows generated by each algorithm for every word in the LabMT dataset in the online appendix (http://compstorylab.org/shocklets/all_word_plots/). A review of figures in the online appendix corresponding with words annotated in Figs. 11 and 12 provides evidence that an anomaly detection algorithm, such as ADV, may not necessarily capture the sane dynamics as does STAR. We include selected panels of these figures in Appendix 3, displaying words corresponding with some peaks of the weighted shock and spike indicator functions. (We hasten to note that this of course does not preclude the possibility that anomaly detection algorithms might indicate dynamics that are not captured by STAR.)

2.3 Typology of local mechanistic dynamics

To further understand divergent dynamic behavior in word rank time series, we analyze regions of these time series for which Eq. (15) is satisfied—that is, where the value of the shock indicator function is greater than the sensitivity parameter. We focus on shock-like dynamics since these dynamics qualitatively describe aggregate social focusing and subsequent de-focusing of attention mediated by the algorithmic substrate of the Twitter platform.

We extract shock segments from the time series of all words that made it into the top $j = 20$ ranked shock indicator functions at least once. Since shocks exist on a wide variety of dynamic ranges and timescales, we normalize all extracted shock segments to lie on the time range $t_{\text{shock}} \in [0, 1]$ and have (spatial) mean zero and variance unity. Shocks have a focal point about their maxima by definition, but in the context of stochastic time series (as considered here), the observed maximum of the time series may not be the "true" maximum of the hypothesized underlying deterministic dynamics. Shock points—hypothesized deterministic maxima—of the extracted shock segments were thus determined by two methods: The maxima of the within-window time series,

$$t_1^* = \underset{t_{\text{shock}} \in [0,1]}{\arg\max}\ x_{t_{\text{shock}}}; \tag{19}$$

and the maxima of the time series's shock indicator function,

$$t_2^* = \underset{t_{\text{shock}} \in [0,1]}{\arg\max}\ C_{\mathcal{K}^{(S)}}(t_{\text{shock}}). \tag{20}$$

We then computed empirical probability density functions of t_1^* and t_2^* across all words in the LabMT dataset. While the empirical distribution of t_1^* is uni-modal, the corresponding empirical distribution of t_2^* demonstrated clear bi-modality with peaks in the first and last quartiles of normalized time. To better characterize these maximum *a posteriori* (MAP)

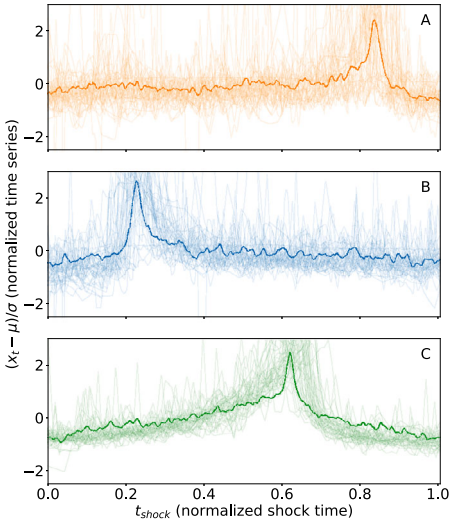

Figure 13 Social dynamics in cusp segments. Extracted cusp Extracted shock segments show diverse behavior corresponding to divergent social dynamics. We extract "important" shock segments (those that breach the top $k = 20$ ranked weighted shock indicator at least once during the decade under study) and normalize them as described in Sect. 2. We then find the densities of shock points t_1^*, measured using the maxima of the within-window time series, and alternatively measured using the maxima of the (relative) shock indicator function. We calculate relative maxima of these distributions and spatially-average shock segments whose maxima were closest to these relative maxima; we display these mean shock segments along with sample shock segments that are close to these mean shock segments in norm. We introduce a classification scheme for shock dynamics: Type I (panel (**A**)) dynamics are those that display slow buildup and fast relaxation; Type II (panel (**B**)) dynamics, conversely, display fast (shock-like) buildup and slow relaxation; and Type III (panel (**C**)) dynamics are relatively symmetric. Overall, we find that Type III dynamics are most common (40.9%) among words that breach the top $k = 20$ ranked weighted shock indicator function, while Type II are second-most common (36.4%), followed by Type I (22.7%)

estimates, we sample those shock segments x_t the maxima of which are temporally-close to the MAPs and calculate spatial means of these samples,

$$\langle x_{t_{\text{shock}}} \rangle_n = \frac{1}{|\mathcal{M}|} \sum_{n \in \mathcal{M}} x^{(n)}_{t_{\text{shock}}}, \tag{21}$$

where

$$.\mathcal{M} = \left\{ n : \left| \arg\max_{t_{\text{shock}} \in [0,1]} x^{(n)}_{t_{\text{shock}}} - t^* \right| < \varepsilon \right\}. \tag{22}$$

The number ε is a small value which we set here to $\varepsilon = 10/503$.[f] We plot these curves in Fig. 13. Shock segments that are close in spatial norm to the $\langle x_{t_{\text{shock}}} \rangle_n$—that is, shock segments $x_{t_{\text{shock}}}$ that satisfy

$$\left\| x_{t_{\text{shock}}} - \langle x_{t_{\text{shock}}} \rangle_n \right\|_1 \leq F^{\leftarrow}_{\|x_s - \langle x_{t_{\text{shock}}} \rangle_n \|_1}(0.01), \tag{23}$$

where $F^{\leftarrow}_Z(q)$ is the quantile function of the random variable Z—are plotted in thinner curves. From this process, three distinct classes of shock segments emerge, corresponding with the three relative maxima of the shock point distributions outlined above:

- *Type I*: exhibiting a slow buildup (anticipation) followed by a fast relaxation;
- *Type II*: with a correspondingly short buildup (shock) followed by a slow relaxation;
- *Type III*: exhibiting a relatively symmetric shape.

Words corresponding to these classes of shock segments differ in semantic context. Type I dynamics are related to known and anticipated societal and political events and subjects, such as:

- "hampshire" and "republican", concerning U.S. presidential primaries and general elections,
- "labor", "labour", and "conservatives", likely concerning U.K. general elections,

Table 1 Words for which at least one shock segment was close in norm to a spatial mean shock segment as detailed in Sect. 2. We display the distributions of "shock points"—hypothesized deterministic maxima of the noisy mechanistically-generated time series—in Fig. 13. Every word may also have several "shock points" where each point could corresponds to a different shock dynamics due to the way each word is used throughout its life span on the platform, hence a few of these examples (e.g. rumble, anonymity, #nowplaying) appear in multiple categories

Classification	Shock shape	Words
Type I	Slow buildup, fast relaxation	rumble, veterans, dusty, labour, scattered, hampshire, #tinychat, elected, ballot, selection, labor, entering, beam, phenomenon, voters, mamma, anonymity, republican, #nowplaying, indictment, wages, conservatives, pulse, knee, grammy, essays, #tcot, kentucky, fml, netherlands, jingle, valid, whitman, syracuse, dems, deposit, bail, tomb, walker, reader
Type II	Fast buildup, slow relaxation	xbox, chained, yale, bombing, holocaust, connecticut, #tinychat, civilian, jill, turkish, tsunami, ferry, #letsbehonest, beam, agreement, riley, ethics, phenomenon, harriet, privacy, israeli, #nowplaying, gun, dub, pulse, killings, herman, enormous, fbi, dmc, searched, norman, joan, affected, arthur, sandra, radiation, army, walker, reader,
Type III	Roughly symmetric	rumble, memorial, sleigh, veterans, costumes, greeks, britney, separated, father's, shark, grammys, labor, costume, x-mas, bunny, commonwealth, clause, olympics, olympic, daylight, cyber, wrapping, rudolph, drowned, re-election

- "voter", "elected", and "ballot", concerning voting in general, and
- "grammy", the music awards show.

To contrast, Type II (shock-like) dynamics describe events that are partially- or entirely-unexpected, often in the context of national or international crises, such as:

- "tsunami" and "radiation", relating to theFukushima Daichii tsunami and nuclear meltdown,
- "bombing", "gun", "pulse", "killings", and "connecticut", concerning acts of violence and mass shootings, in particular the Sandy Hook elementary school shooting in the United States;
- "jill" (Jill Stein, a 2016 U.S. presidential election competitor), "ethics", and "fbi", pertaining to surprising events surrounding the 2016 U.S. presidential election, and
- "turkish", "army", "israeli", "civilian", and "holocaust", concerning international protests, conflicts, and coups.

Type III dynamics are associated with anticipated events that typically re-occur and are discussed substantially after their passing, such as

- "sleigh", "x-mas", "wrapping", "rudolph", "memorial", "costumes", "costume", "veterans", and "bunny", having to do with major holidays, and
- "olympic" and "olympics", relating to the Olympic games.

We give a full list of words satisfying the criteria given in Eqs. (22) and (23) in Table 1. We note that, though the above discussion defines and distinguishes three fundamental signatures of word rank shock segments, these classes are only the MAP estimates of the true distributions of shock segments, our empirical observations of which are displayed as histograms in Fig. 13; there is an effective continuum of dynamics that is richer, but more complicated, than our parsimonious description here.

3 Discussion

We have introduced a nonparametric pattern detection method, termed the discrete shocklet transform (DST) for its particular application in extracting shock- and shock-like

dynamics from noisy time series, and demonstrated its particular suitability for analysis of sociotechnical data. Though extracted social dynamics display a continuum of behaviors, we have shown that maximizing *a posteriori* estimates of shock likelihood results in three distinct classes of dynamics: anticipatory dynamics with long buildups and quick relaxations, such as political contests (Type I); "surprising" events with fast (shock-like) buildups and long relaxation times, examples of which are acts of violence, natural disasters, and mass shootings (Type II); and quasi-symmetric dynamics, corresponding with anticipated and talked-about events such as holidays and major sporting events (Type III). We analyzed the most "important" shock-like dynamics—those words that were one of the top-20 most significant at least once during the decade of study—and found that Type III dynamics were the most common among these words (40.9%) followed by Type II (36.4%) and Type I (22.7%). We then showcased the discrete shocklet transform's effectiveness in extracting coherent intertemporal narratives from word usage data on the social microblog Twitter, developing a graphical methodology for examining meaningful fluctuations in word—and hence topic—popularity. We used this methodology to create document-free nonparametric topic models, represented by pruned networks based on shock indicator similarity between two words and defining topics using the networks' community structures. This construction, while retaining artifacts from its construction using intrinsically-temporal data, presents topics possessing qualitatively sensible semantic structure.

There are several areas in which future work could improve on and extend that presented here. Though we have shown that the discrete shocklet transform is a useful tool in understanding non-stationary local behavior when applied to a variety of sociotechnical time series, there is reason to suspect that one can generalize this method to essentially any kind of noisy time series in which it can be hypothesized that mechanistic local dynamics contribute a substantial component to the overall signal. In addition, the DST suffers from noncausality, as do all convolution or frequency-space transforms. In order to compute an accurate transformed signal at time t, information about time $t + \tau$ must be known to avoid edge effects or spectral effects such as ringing. In practice this may not be an impediment to the DST's usage, since: empirically the transform still finds "important" local dynamics, as shown in Fig. 11 near the very beginning (the words "occupy" and "slumdog" are annotated) and the end (the words "stormy" and "cohen" are annotated) of time studied. Furthermore, when used with more frequently-sampled data the lag needed to avoid edge effects may have decreasing length relative to the longer timescale over which users interact with the data. However, to avoid the problem of edge effects entirely, it may be possible to train a supervised learning algorithm to learn the output of the DST at time t using only past (and possibly present) data. The DST could also serve as a useful counterpart to phrase- and sentence-tracking algorithms such as MemeTracker [93, 94]. Instead of applying the DST to time series of simple words, one could apply it to arbitrary n-grams (including whole sentences) or sentence structure pattern matches to uncover frequency of usage of verb tenses, passive/active voice construction, and other higher-order natural language constructs. Other work could apply the DST to more and different natural language data sources or other sociotechnical time series, such as asset prices, economic indicators, and election polls.

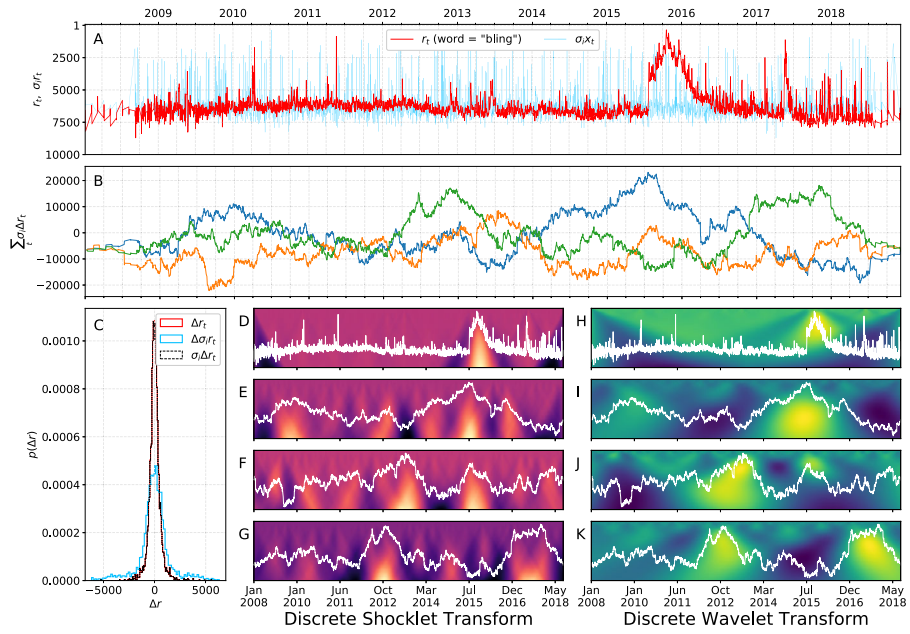

Figure 14 Intricate dynamics of sociotechnical time series. Sociotechnical time series can display intricate dynamics and extended periods of anomalous behavior. The red curve shows the time series of the ranks down from top of the word "bling" on Twitter. Until 2015/10/31, the time series presents as random fluctuation about a steady trend that is nearly indistinguishable from zero. However, the series then displays a large fluctuation, increases rapidly, and then decays slowly after a sharp peak. The underlying mechanism for these dynamics was the release of a popular song titled "Hotline Bling" by a musician known as "Drake". Returns $\Delta r_t = r_{t+1} - r_t$ are calculated and their histogram is displayed in panel (**C**). To demonstrate the qualitative difference of the "bling" time series from other time series with an identical returns distribution, elements of the symmetric group $\sigma_i \in \mathcal{S}_T$ are applied to the returns of the original series, $\Delta r_t \mapsto \Delta r_{\sigma_i t}$, and the resultant noise is integrated and plotted as $r_{\sigma_i t} = \sum_{t' \leq t} \Delta r_{\sigma_i t}$. The bottom-left panel (**C**) displays time-decoupled probability distributions of the returns of the plotted time series. The distributions of Δr_i and $\sigma \Delta r_i$ are identical, as they should be, but the integrated series have entirely different spectral behavior and dynamic ranges. Panels (**D**)–(**G**) display the discrete shocklet transform of the original series and the random walks $\sum_{t' \leq t} \Delta r_{\sigma_i t}$, showing the responsiveness of the DST to nonstationary local dynamics and its insensitivity to dynamic range. The right-most column of panels (**H**)–(**K**) displays the discrete wavelet transform of the original series demonstrating its comparatively less-sensitive nature to local anomalous dynamics

Appendix 1: Statistical details

In this appendix we will outline some statistical details of the DST and STAR algorithm that are not necessary for a qualitative understanding of them, but could be useful for more in-depth understanding or efforts to generalize them.

We first give an illustrative example of how a sociotechnical time series can differ substantially from two null models of time series that have some similar statistical properties, displayed in Fig. 14 (a more information-rich version of Fig. 5, displayed in the main body), panels (A) and (B). In panel (A), we display an example sociotechnical time series in the red curve, usage rank of the word "bling" within the LabMT subset of words on Twitter (denoted by r_t), and σr_t, a randomly shuffled version of this time series. We denote $\sigma \in \mathcal{S}_T$, the symmetric group on T elements, and draw σ from the uniform distribution over $\mathcal{S}_T$. It is immediately apparent that the structure of r_t and σr_t are radically different in autocorrelation (both in levels and differences) and we do not investigate this admittedly-naïve null model any further.

We next consider a random walk null model constructed as follows: first differencing r_t to obtain $\Delta r_t = r_t - r_{t-1}$, we apply random elements $\sigma_i \in \mathcal{S}_T$ and integrate, displaying the resulting $r_{\sigma_i t} = \sum_{t' \leq t} \sigma_i \Delta r_t$ in panel (C) of Fig. 14. Visual inspection (i.e., the "eye test") also demonstrates that these time series do not replicate the behavior displayed by the original r_t; they become negative, have a dynamic range that is almost an order of magnitude larger, and are more highly autocorrelated. We contrast the results of the DST on r_t and draws from this random walk null model in panels (D)–(G) of Fig. 14. In panel (D) we display the DST of r_t, while in panels (E)–(G) we display the DST of three random $\sigma_i r_t$. The DSTs of the draws from the random walk model are more irregular that the DST of r_t, displaying many time-domain fluctuations between large positive values and large negative values. In contrast, the DST of r_t is relatively constant except near August of 2015, where it exhibits a large positive fluctuation across a wide range of W. The underlying dynamics for this fluctuation were driven by the release of a popular song called "Hotline Bling" on July 31st, 2015.

As a couterpoint to the DST, we computed the discrete wavelet transform (DWT) of r_t and the same $\sigma_i r_t$. We computed the wavelet transform using the Ricker wavelet,

$$\psi(\tau, W) = \frac{2}{\sqrt{3W\pi^{1/2}}}\left[1 - \left(\frac{\tau}{W}\right)^2\right] e^{-\tau^2/(2W^2)}. \tag{24}$$

We chose to compare the DST with the DWT because these transforms are very similar in many respects: they both depend on two parameters (a location parameter τ and a scale parameter W); they both output a matrix of shape $T \times N_W$ (N_W rows, one for each value W, and T columns, one for each value of τ). There are some key difference between these transforms, however. The "kernels" of the wavelet transform—the kernels—have unique properties not shared by our shock-like kernels: wavelets $\psi(t)$ are defined on all of $\mathbb{R}$, satisfy $\lim_{t \to \pm\infty} \psi(t) = 0$, and are orthonormal. Our shock-like kernels do not satisfy any of these properties; they are defined on a finite interval $[-W/2, W/2]$, do not vanish at the endpoints of this interval, and are not orthogonal functions. Hence, differences in the DST and DWT of a time series are due primarily to the choice of convolution function—shock-like kernel in the case of the DST and wavelet in the case of the DWT. We display the DWT of r_t and the same $\sigma_i r_t$ in panels (H)–(K) of Fig. 14. Comparing these transforms with the DSTs displayed in panels (D)–(G), we see that the DST has increased time-localization over the DWT in time intervals during which the time series exhibit shock-like dynamics.

As we note in Sect. 2.1 (there when comparing STAR to Twitter's ADV anomaly detection algorithm), this observation should not be construed as equivalent to the statement that the DST is in some way superior to the DWT or should supersede the DWT for general time series processing tasks; rather, it is evidence that the DST is a superior transform than the DWT for the purpose of finding shock-like dynamics in sociotechnical time series—a task for which it was designed and the DWT was not.

We finally note an analytical property of the DST that, while likely not useful in practice, is a fact that should be recorded and may be useful in constructing theoretical extensions of the DST. The DST is defined in Eq. (11), which we record here for ease in reference:

$$C_{\mathcal{K}^{(\cdot)}}(t, W|\theta) = \sum_{-\infty}^{\infty} x(t+\tau)\mathcal{K}^{(\cdot)}(\tau|W, \theta), \tag{25}$$

defined for each t. The function $\mathcal{K}^{(\cdot)}$ is the shock kernel that is non-zero on $\tau \in [-W/2 + t, W/2 + t]$. For $t \in [-T, T]$, this can be rewritten equivalently as

$$\mathbf{C}_{\mathcal{K}^{(\cdot)}}(W|\theta) = \mathbf{K}(W|\theta)\mathbf{x}, \tag{26}$$

where $\mathbf{K}(W|\theta)$ is a $(2T + 1) \times (2T + 1)$ W-diagonal matrix, $\mathbf{C}_{\mathcal{K}^{(\cdot)}}(W|\theta)$ is the Wth row of the cusplet transform matrix, and $\mathbf{x}$ is the entire time series $x(t)$ considered as a vector in $\mathbb{R}^{2T+1}$. The matrix $\mathbf{K}(W|\theta)$ is just the convolution matrix corresponding to the cross-correlation operation with $\mathcal{K}^{(\cdot)}$. If $\mathbf{K}(W|\theta)$ is invertible, then it is clear that

$$\mathbf{x} = \mathbf{K}(W|\theta)^{-1}\mathbf{C}_{\mathcal{K}^{(\cdot)}}(W|\theta), \tag{27}$$

for any $1 < W < T$ and hence also

$$\mathbf{x} = \frac{1}{N_W}\sum_{W}\mathbf{K}(W|\theta)^{-1}\mathbf{C}_{\mathcal{K}^{(\cdot)}}(W|\theta). \tag{28}$$

This is an inversion formula similar to the inversion formulae of overcomplete transforms such as the DWT and discrete chirplet transform.

When $T \to \infty$ (that is, when the signal $x(t)$ is turned on in the infinite past and continues into the infinite future), this equation becomes the formal operator equation

$$C_{\mathcal{K}^{(\cdot)}}(t, W|\theta) = \mathbf{K}(W|\theta)\big[x(t)\big], \tag{29}$$

and hence (as long as the operator inverses are well-defined),

$$x(t) = \frac{1}{N_W}\sum_{W}\mathbf{K}(W|\theta)^{-1}\big[\mathrm{C}_{\mathcal{K}^{(\cdot)}}(t, W|\theta)\big]. \tag{30}$$

These inversion formulae are, in our estimation, of relatively little utility in practical application. Whereas inverting a wavelet transform is a common task—it may be desirable to decompress an image that is initially compressed using the JPEG 2000 algorithm, which uses the wavelet transform for compact representation of the image—we estimate the probability of being presented with some arbitrary shocklet transform and needing to recover the original signal from it to be quite low; the shocklet transform is designed to amplify features of signals to which we already have access, not to recreate time-domain signals from their representations in other domains.

Appendix 2: Document-free topic networks

An important application of the DST is the partial recovery of context- or document-dependent information from aggregated time series data. In natural language processing, many models of human language are statistical in nature and require original documents from which to infer values of parameters and perform estimation [95, 96]. However, such information can be both expensive to purchase and require a large amount of physical storage space. For example, the tweet corpus from which the labMT rank dataset used throughout this paper was originally derived is not inexpensive and requires approximately 55 TB of disk space for storage.[g] In contrast, the dataset used here is derived from

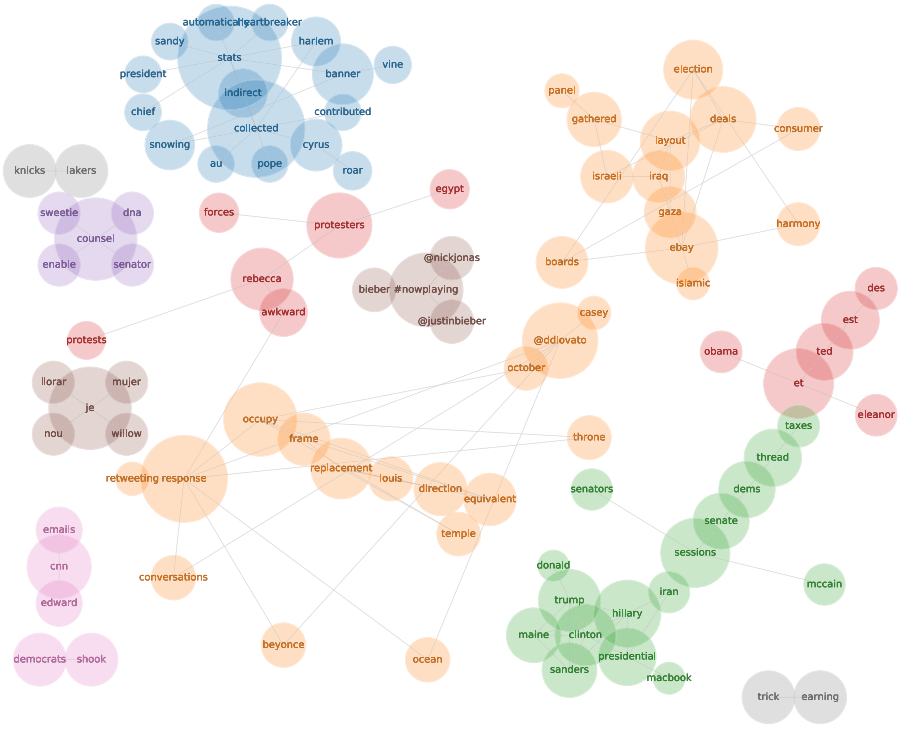

Figure 15 Topic network inferred from weighted shock indicator functions. At each point in time, words are ranked according to the value of their weighted shock indicator function and the top k words are taken and linked pairwise for an upper bound of $\binom{k}{2}$ additional edges in the network; if the edge between words i and j already exists, the weight of the edge is incremented. The edge weight increment at time t is given by $w_{ij,t} = \frac{R_{i,t}+R_{j,t}}{2}$, the average of the weighted shock indicator for words i and j, with the total edge weight thus given by $w_{ij} = \sum_t w_{ij,t}$. After initial construction, the backbone of the network is extracted using the method of Serrano *et al.* [97]. The network is pruned further by retaining only those nodes i, j and edges e_{ij} for which w_{ij} is above the pth percentile of all edge weights in the backboned network. The network displayed here is constructed by setting $k = 20$ and $p = 50$, where size of the node indicates normalized page rank. Topics are associated with distinct communities, found using the modularity algorithm of Clauset *et al.* [98]

the freely-available LabMT word set and is less than 400 MB in size. If topics of relatively comparable quality can be extracted from this smaller and less expensive dataset, the potential utility to the scientific community at large, could be high.

We demonstrate that a reasonable topic model for Twitter during the time period of study can be inferred from the panel of rank time series alone. This is accomplished via a multi-step meta-algorithm. First, the weighted Shock Indicator Function R_i is calculated for each word i. At each point in time t, words are sorted by their respective shock indicator functions. At time step t, the top k words are taken and linked pairwise for an upper bound of $\binom{k}{2}$ additional edges in the network; if an edge already exists between word i and j, it is incremented by the mean of the words' respective weighted Shock Indicator Function $\frac{R_i+R_j}{2}$. Performing this process for all time periods results in a weighted network of related words. The weights $w_{ij} = \sum_t \frac{R_{i,t}+R_{j,t}}{2}$ are large when the value of a word's weighted shock indicator function is large or when a word is frequently in the top k, even if it is never near the top. The resulting network can be large; to reduce its size, its backbone is extracted using the method of Serrano *et al.* [97] and further pruned by retaining only those nodes and edges for which the corresponding edge weights are at or above the pth percentile

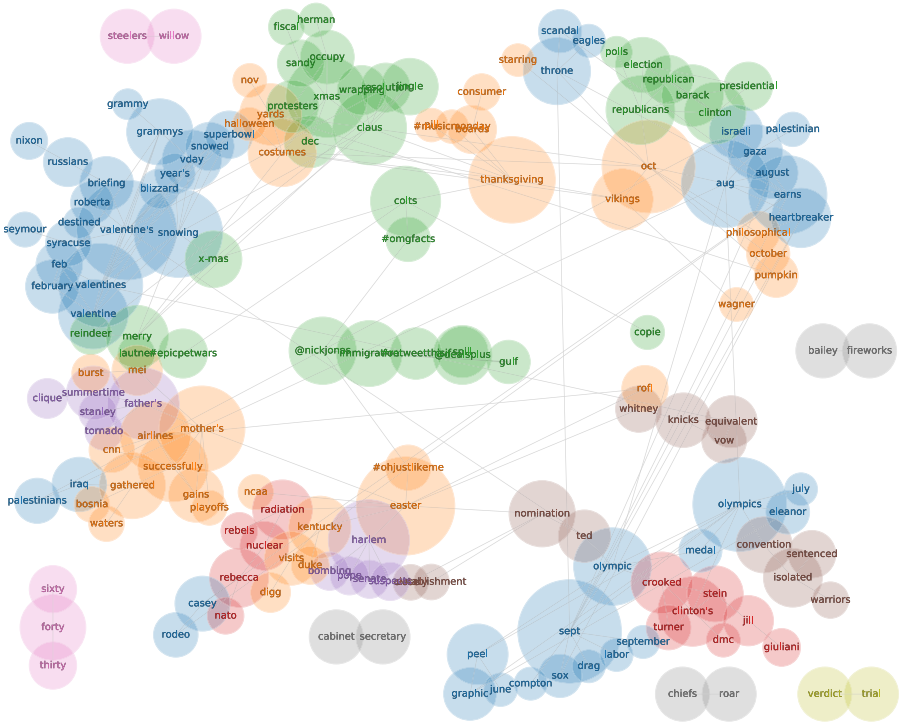

Figure 16 Topic network inferred from weighted spike indicator functions. At each point in time, words are ranked according to the value of their weighted spike indicator function and the top k words are taken and linked pairwise for an upper bound of $\binom{k}{2}$ additional edges in the network; if the edge between words i and j already exists, the weight of the edge is incremented. The edge weight increment at time t is given by $w_{ij,t} = \frac{R_{i,t}+R_{j,t}}{2}$, the average of the weighted spike indicator for words i and j, with the total edge weight thus given by $w_{ij} = \sum_t w_{ij,t}$. After initial construction, the backbone of the network is extracted using the method of Serrano *et al.* [97]. The network is pruned further by retaining only those nodes i, j and edges e_{ij} for which w_{ij} is above the pth percentile of all edge weights in the backboned network. The network displayed here is constructed by setting $k = 20$ and $p = 50$, where size of the node indicates normalized page rank. Topics are associated with distinct communities, found using the modularity algorithm of Clauset *et al.* [98]

of all weights in the backboned network. Topics are associated with communities in the resulting pruned networks, found using the modularity algorithm of Clauset *et al.* [98].

Figure 15 and Fig. 16 display the result of this procedure for $k = 20$ and $p = 50$. Unique communities (topics) are indicated by node color. In the co-shock network (Fig. 15), topics include, among others:

- Winter holidays and events ("valentines", "superbowl", "vday", …);
- U.S. presidential elections ("republicans", "barack", "clinton", "presidential", …);
- Events surrounding the 2016 U.S. presidential election in particular ("clinton's", "crooked", "giuliani", "jill", "stein", …);

while the co-shock network displays topics pertaining to:

- popular culture and music ("bieber", "#nowplaying", "@nickjonas", "@justinbieber");
- U.S. domestic politics ("clinton", "hillary", "trump", "sanders", "iran", "sessions", …);
- and conflict in the Middle East ("gaza", "iraq", "israeli", "gathered")

The predominance of U.S. politics at the exclusion of politics of other nations is likely because the labMT dataset contains predominantly English words.

Appendix 3: STAR and ADV comparison figures

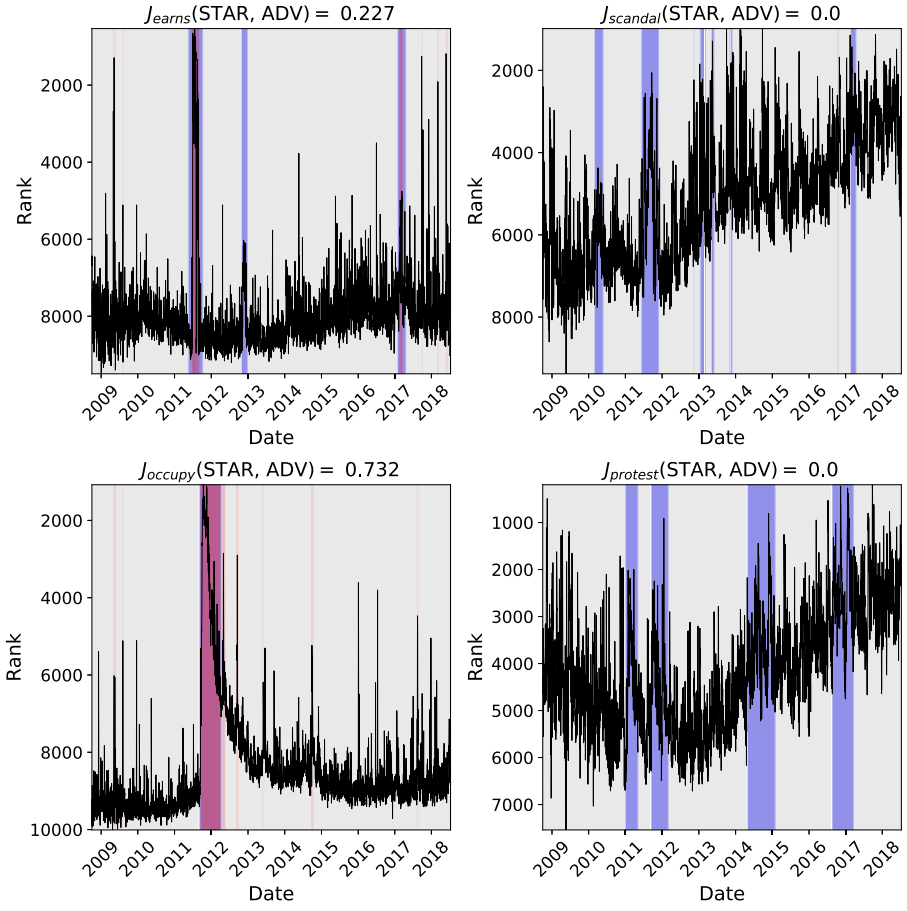

Figure 17 Occupy Wall Street. Comparison of STAR and ADV indicator windows for some words surrounding the "Occupy Wall Street" movement during 2010

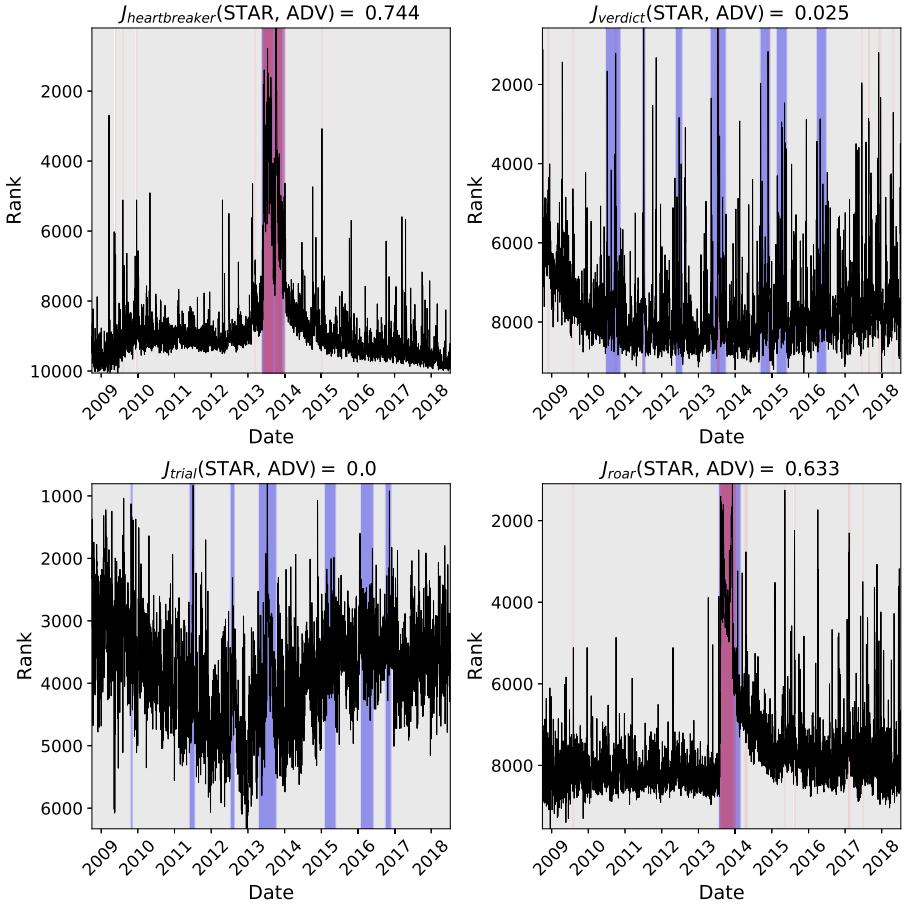

Figure 18 Heartbreaker. Comparison of STAR and ADV indicator windows for some words surrounding popular events (the release of a song called "Heartbreaker" by Justin Bieber and "Roar" by Katy Perry) and criminal justice-related events (the trial and acquittal of George Zimmerman)

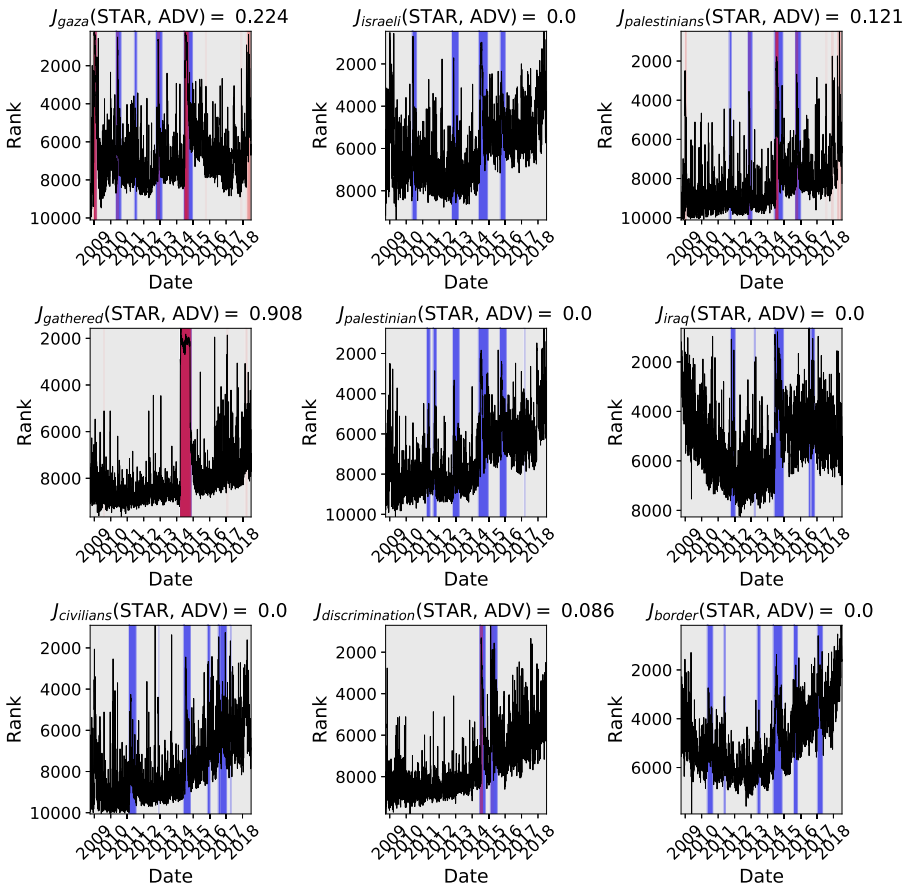

Figure 19 Gaza conflict. Comparison of STAR and ADV indicator windows for some words surrounding the Gaza conflict of 2014

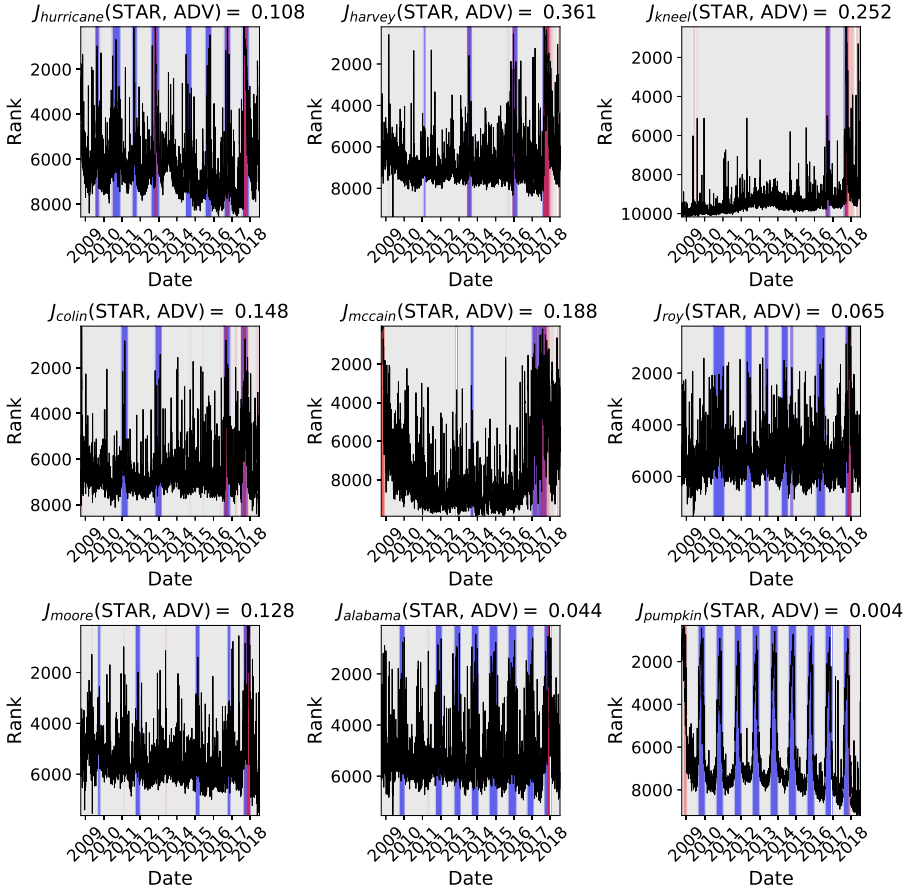

Figure 20 Hurricanes. Comparison of STAR and ADV indicator windows for some words surrounding the autumn of 2017, including Hurricane Harvey, Colin Kaepernick's kneeling protests, John McCain, the electoral campaign of Roy Moore in the U.S. state of Alabama, and pumpkins (a traditional gourd symbolic of autumn in the U.S.)

Acknowledgements
The authors acknowledge the computing resources provided by the Vermont Advanced Computing Core and financial support from the Massachusetts Mutual Life Insurance Company, and are grateful for web hosting assistance from Kelly Gothard and useful conversations with Jane Adams and Colin Van Oort.

Funding
The authors acknowledge financial support from NSF Big Data Grant #1447634 and MassMutual Life Insurance.

Abbreviations
ADV, anomaly detection vector, Twitter's anomaly detection algorithm; DST, Discrete Shocklet Transform, the analytical method at the core of this article; STAR, Shocklet Transform and Ranking algorithm, method that uses the DST to provide qualitative, shape-based similarity search in multivariate time series; DWT, Discrete Wavelet Transform, a transform that performs a task similar to, yet fundamentally different from, that performed by the DST; LabMT, Lab Mechanical Turk dataset, set of over 10,000 words used here and in multiple studies of human language and behavior; MAP, Maximum *a posteriori*, estimated maximum of a posterior distribution; WSIF, weighted shock indicator function, one of the outputs of the STAR algorithm.

Availability of data and materials
The datasets analysed during the current study and our source code are available in the https://gitlab.com/compstorylab/discrete-shocklet-transform repository.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
DRD, DK, CMD, and PSD conceived of the idea; DRD and DK developed the theory; DRD, TA, MVA, and JRM analyzed data; DRD, TA, MVA, JRM, and PSD wrote the paper; DRD, TA, MVA, JRM, CMD, and PSD edited the paper. All authors read and approved the final manuscript.

Author details
[1]Computational Story Lab, Vermont Complex Systems Center, University of Vermont, Burlington, United States. [2]Department of Mathematics & Statistics, University of Vermont, Burlington, United States. [3]Department of Computer Science, University of Vermont, Burlington, United States. [4]Zillow Group, Seattle, United States.

Endnotes
[a] Python implementations of the DST and STAR algorithms are located at this git repository: https://gitlab.com/compstorylab/discrete-shocklet-transform.
[b] Bayesian classification algorithms can perform classification based only on prior information, but this is also not similar to the STAR algorithm, since the STAR algorithm is a maximum-likelihood method that by definition requires at least one time series to operate.
[c] https://github.com/twitter/AnomalyDetection.
[d] Data is available at https://fred.stlouisfed.org/series/RECPROUSM156N.
[e] http://compstorylab.org/shocklets/.
[f] This value comes from an arbitrary but small number of indices (five) we allow a shock segment to vary (±) about the index of the MAP estimate of the distributions of shock points, each of which can be considered as multinomial distributions supported on a 503-dimensional vector space. The number 503 is the dimension of each shock segment after time normalization since the longest original shock segment in the labMT dataset was 503 days.
[g] The dataset is available for purchase from Twitter at http://support.gnip.com/apis/firehose/overview.html. The on-disk memory statistic is the result of `du -h <dirname> | tail -n 1` on the authors' computing cluster and so may vary by machine or storage system.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 28 June 2019 Accepted: 13 January 2020 Published online: 07 February 2020

References
1. Chaovalit P, Gangopadhyay A, Karabatis G, Chen Z (2011) Discrete wavelet transform-based time series analysis and mining. ACM Comput Surv (CSUR) 43(2):6
2. Yeh C-CM, Kavantzas N, Keogh E (2017) Matrix profile vi: meaningful multidimensional motif discovery. In: 2017 IEEE international conference on data mining (ICDM). IEEE Press, New York, pp 565–574
3. Zhu Y, Imamura M, Nikovski D, Keogh E (2018) Introducing time series chains: a new primitive for time series data mining. Knowl Inf Syst: 1–27
4. Struzik ZR, Siebes AP (2002) Wavelet transform based multifractal formalism in outlier detection and localisation for financial time series. Phys A, Stat Mech Appl 309(3–4):388–402
5. Popivanov I, Miller RJ (2002) Similarity search over time-series data using wavelets. In: Proceedings 18th international conference on data engineering. IEEE Press, New York, pp 212–221
6. Lau K-M, Weng H (1995) Climate signal detection using wavelet transform: how to make a time series sing. Bull Am Meteorol Soc 76(12):2391–2402

7. Whitcher B, Byers SD, Guttorp P, Percival DB (2002) Testing for homogeneity of variance in time series: long memory, wavelets, and the Nile river. Water Resour Res 38(5)
8. Benítez R, Bolós V, Ramírez M (2010) A wavelet-based tool for studying non-periodicity. Comput Math Appl 60(3):634–641
9. Mann S, Haykin S (1991) The chirplet transform: a generalization of Gabor's logon transform. In: Vision interface, vol 91, pp 205–212
10. Wang G, Xia X-G, Root BT, Chen VC (2002) Moving target detection in over-the-horizon radar using adaptive chirplet transform. In: Proceedings of the 2002 IEEE radar conference (IEEE cat. no. 02CH37322). IEEE Press, New York, pp 77–84
11. Spanos P, Giaralis A, Politis N (2007) Time–frequency representation of earthquake accelerograms and inelastic structural response records using the adaptive chirplet decomposition and empirical mode decomposition. Soil Dyn Earthq Eng 27(7):675–689
12. Taebi A, Mansy H (2016) Effect of noise on time-frequency analysis of vibrocardiographic signals. J Bioeng & Biomed Sci 6(4)
13. Page E (1955) A test for a change in a parameter occurring at an unknown point. Biometrika 42(3/4):523–527
14. Mallat S, Hwang WL (1992) Singularity detection and processing with wavelets. IEEE Trans Inf Theory 38(2):617–643
15. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS ONE 6(12):26752
16. Li Q, Shah S, Thomas M, Anderson K, Liu X, Nourbakhsh A, Fang R (2017) How much data do you need? Twitter decahose data analysis
17. Reagan AJ, Danforth CM, Tivnan B, Williams JR, Dodds PS (2017) Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. EPJ Data Sci 6(1):28
18. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ (2017) Forecasting the onset and course of mental illness with Twitter data. Sci Rep 7(1):13006
19. Frank MR, Mitchell L, Dodds PS, Danforth CM (2013) Happiness and the patterns of life: a study of geolocated tweets. Sci Rep 3:2625
20. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM (2013) The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. PLoS ONE 8(5):64417
21. Lemahieu R, Van Canneyt S, De Boom C, Dhoedt B (2015) Optimizing the popularity of Twitter messages through user categories. In: 2015 IEEE international conference on data mining workshop (ICDMW). IEEE Press, New York, pp 1396–1401
22. Wu F, Huberman BA (2007) Novelty and collective attention. Proc Natl Acad Sci 104(45):17599–17601
23. Candia C, Jara-Figueroa C, Rodriguez-Sickert C, Barabási A-L, Hidalgo CA (2019) The universal decay of collective memory and attention. Nat Hum Behav 3(1):82
24. Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. Proc Natl Acad Sci 105(41):15649–15653
25. Lorenz-Spreen P, Mønsted BM, Hövel P, Lehmann S (2019) Accelerating dynamics of collective attention. Nat Commun 10(1):1759
26. De Domenico M, Altmann EG (2019) Unraveling the origin of social bursts in collective attention. arXiv preprint. arXiv:1903.06588
27. Ierley G, Kostinski A (2019) A universal rank-order transform to extract signals from noisy data. arXiv preprint. arXiv:1906.08729
28. Nakamoto S, et al. (2008) Bitcoin: a peer-to-peer electronic cash system
29. Al Shehhi A, Oudah M, Aung Z (2014) Investigating factors behind choosing a cryptocurrency. In: 2014 IEEE international conference on industrial engineering and engineering management. IEEE Press, New York, pp 1443–1447
30. Lin J, Keogh E, Wei L, Lonardi S (2007) Experiencing sax: a novel symbolic representation of time series. Data Min Knowl Discov 15(2):107–144
31. Yang K, Shahabi C (2007) An efficient k nearest neighbor search for multivariate time series. Inf Comput 205(1):65–98
32. Kale DC, Gong D, Che Z, Liu Y, Medioni G, Wetzel R, Ross P (2014) An examination of multivariate time series hashing with applications to health care. In: 2014 IEEE international conference on data mining. IEEE Press, New York, pp 260–269
33. Driemel A, Silvestri F (2017) Locality-sensitive hashing of curves. arXiv preprint. arXiv:1703.04040
34. Keogh EJ, Pazzani MJ (2000) A simple dimensionality reduction technique for fast similarity search in large time series databases. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, pp 122–133
35. Wu Y-L, Agrawal D, El Abbadi A (2000) A comparison of dft and dwt based similarity search in time-series databases. In: Proceedings of the ninth international conference on information and knowledge management. ACM, New York, pp 488–495
36. Chan F-P, Fu A-C, Yu C (2003) Haar wavelets for efficient similarity search of time-series: with and without time warping. IEEE Trans Knowl Data Eng 15(3):686–705
37. Ratanamahatana C, Keogh E, Bagnall AJ, Lonardi S (2005) A novel bit level time series representation with implication of similarity search and clustering. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, pp 771–777
38. Keogh E, Lin J, Fu A (2005) Hot sax: efficiently finding the most unusual time series subsequence. In: Fifth IEEE international conference on data mining (ICDM'05). IEEE Press, New York, p 8
39. Yeh C-CM, Zhu Y, Ulanova L, Begum N, Ding Y, Dau HA, Silva DF, Mueen A, Keogh E (2016) Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In: 2016 IEEE 16th international conference on data mining (ICDM). IEEE Press, New York, pp 1317–1322
40. Eastman JR, Fulk M (1993) Long sequence time series evaluation using standardized principal components. Photogramm Eng Remote Sens 59(6)
41. Harris D (1997) Principal components analysis of cointegrated time series. Econom Theory 13(4):529–557
42. Lansangan JRG, Barrios EB (2009) Principal components analysis of nonstationary time series data. Stat Comput 19(2):173

43. Mueen A, Viswanathan K, Gupta C, Keogh E (2017) The fastest similarity search algorithm for time series subsequences under Euclidean distance
44. Seref O, Fan Y-J, Chaovalitwongse WA (2013) Mathematical programming formulations and algorithms for discrete k-median clustering of time-series data. INFORMS J Comput 26(1):160–172
45. Vlachos M, Lin J, Keogh E, Gunopulos D (2003) A wavelet-based anytime algorithm for k-means clustering of time series. In: Proc. Workshop on clustering high dimensionality data and its applications. Citeseer
46. Goutte C, Toft P, Rostrup E, Nielsen F, Hansen LK (1999) On clustering fmri time series. NeuroImage 9(3):298–310
47. Jiang D, Pei J, Zhang A (2003) Dhc: a density-based hierarchical clustering method for time series gene expression data. In: Third IEEE symposium on bioinformatics and bioengineering, 2003. Proceedings. IEEE Press, New York, pp 393–400
48. Rodrigues PP, Gama J, Pedroso JP (2006) Odac: hierarchical clustering of time series data streams. In: Proceedings of the 2006 SIAM international conference on data mining. SIAM, Philadelphia, pp 499–503
49. Rodrigues PP, Gama J, Pedroso J (2008) Hierarchical clustering of time-series data streams. IEEE Trans Knowl Data Eng 20(5):615–627
50. Denton A (2005) Kernel-density-based clustering of time series subsequences using a continuous random-walk noise model. In: Fifth IEEE international conference on data mining (ICDM'05). IEEE Press, New York, p 8
51. Birant D, Kut A (2007) St-dbscan: an algorithm for clustering spatial–temporal data. Data Knowl Eng 60(1):208–221
52. Çelik M, Dadaşer-Çelik F, Dokuz AŞ (2011) Anomaly detection in temperature data using dbscan algorithm. In: 2011 international symposium on innovations in intelligent systems and applications. IEEE Press, New York, pp 91–95
53. Kumar M, Patel NR, Woo J (2002) Clustering seasonality patterns in the presence of errors. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 557–563
54. Oates T, Firoiu L, Cohen PR (1999) Clustering time series with hidden Markov models and dynamic time warping. In: Proceedings of the IJCAI-99 workshop on neural, symbolic and reinforcement learning methods for sequence learning, pp 17–21. Citeseer
55. Schreiber T, Schmitz A (1997) Classification of time series data with nonlinear similarity measures. Phys Rev Lett 79(8):1475
56. Kalpakis K, Gada D, Puttagunta V (2001) Distance measures for effective clustering of arima time-series. In: Proceedings 2001 IEEE international conference on data mining. IEEE Press, New York, pp 273–280
57. Bagnall AJ, Janacek GJ (2004) Clustering time series from arma models with clipped data. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 49–58
58. Xiong Y, Yeung D-Y (2004) Time series clustering with arma mixtures. Pattern Recognit 37(8):1675–1689
59. Fröhwirth-Schnatter S, Kaufmann S (2008) Model-based clustering of multiple time series. J Bus Econ Stat 26(1):78–89
60. Keogh E, Lin J (2005) Clustering of time-series subsequences is meaningless: implications for previous and future research. Knowl Inf Syst 8(2):154–177
61. Bagnall A, Lines J, Bostrom A, Large J, Keogh E (2017) The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min Knowl Discov 31(3):606–660
62. Gilbert AC, Kotidis Y, Muthukrishnan S, Strauss M (2001) Surfing wavelets on streams: one-pass summaries for approximate aggregate queries. In: Vldb, vol 1, pp 79–88
63. Ahmad S, Taskaya-Temizel T, Ahmad K (2004) Summarizing time series: learning patterns in 'volatile' series. In: International conference on intelligent data engineering and automated learning. Springer, Berlin, pp 523–532
64. Castillo-Ortega R, Marín N, Sánchez D, Tettamanzi AG (2011) A multi-objective memetic algorithm for the linguistic summarization of time series. In: Proceedings of the 13th annual conference companion on genetic and evolutionary computation. ACM, New York, pp 171–172
65. Castillo Ortega R, Marín N, Sánchez D, Tettamanzi AG (2011) Linguistic summarization of time series data using genetic algorithms. In: EUSFLAT, vol 1. Atlantis Press, pp 416–423
66. Kacprzyk J, Wilbik A, Zadrożny S (2007) Linguistic summarization of time series under different granulation of describing features. In: International conference on rough sets and intelligent systems paradigms. Springer, Berlin, pp 230–240
67. Kacprzyk J, Wilbik A, Zadrożny S (2008) Linguistic summarization of time series using a fuzzy quantifier driven aggregation. Fuzzy Sets Syst 159(12):1485–1499
68. Kacprzyk J, Wilbik A, Zadrożny S (2010) An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation. Int J Intell Syst 25(5):411–439
69. Li L, McCann J, Pollard NS, Faloutsos C (2009) Dynammo: mining and summarization of coevolving sequences with missing values. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 507–516
70. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv (CSUR 41(3):15
71. Gbur G, Visser T, Wolf E (2001) Anomalous behavior of spectra near phase singularities of focused waves. Phys Rev Lett 88(1):013901
72. Plerou V, Gopikrishnan P, Amaral LAN, Gabaix X, Stanley HE (2000) Economic fluctuations and anomalous diffusion. Phys Rev E 62(3):3023
73. Jeon J-H, Tejedor V, Burov S, Barkai E, Selhuber-Unkel C, Berg-Sørensen K, Oddershede L, Metzler R (2011) In vivo anomalous diffusion and weak ergodicity breaking of lipid granules. Phys Rev Lett 106(4):048103
74. Palfrey TR, Prisbrey JE (1997) Anomalous behavior in public goods experiments: how much and why?. Am Econ Rev: 829–846
75. Capra CM, Goeree JK, Gomez R, Holt CA (1999) Anomalous behavior in a traveler's dilemma? Am Econ Rev 89(3):678–690
76. Rosner B (1983) Percentage points for a generalized esd many-outlier procedure. Technometrics 25(2):165–172
77. Vallis O, Hochenbaum J, Kejariwal A (2014) A novel technique for long-term anomaly detection in the cloud. In: 6th {USENIX} workshop on hot topics in cloud computing (HotCloud 14)
78. Laptev N, Amizadeh S, Flint I (2015) Generic and scalable framework for automated time-series anomaly detection. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1939–1947

79. Chan PK, Mahoney MV (2005) Modeling multiple time series for anomaly detection. In: Fifth IEEE international conference on data mining (ICDM'05). IEEE Press, New York, p 8
80. Cheng H, Tan P-N, Potter C, Klooster S (2009) Detection and characterization of anomalies in multivariate time series. In: Proceedings of the 2009 SIAM international conference on data mining. SIAM, Philadelphia, pp 413–424
81. Qiu H, Liu Y, Subrahmanya NA, Li W (2012) Granger causality for time-series anomaly detection. In: 2012 IEEE 12th international conference on data mining. IEEE Press, New York, pp 1074–1079
82. Akouemo HN, Povinelli RJ (2016) Probabilistic anomaly detection in natural gas time series data. Int J Forecast 32(3):948–956
83. Chauvet M (1998) An econometric characterization of business cycle dynamics with factor structure and regime switching. Int Econ Rev: 969–996
84. Dueker M (2005) Dynamic forecasts of qualitative variables: a qual var model of US recessions. J Bus Econ Stat 23(1):96–104
85. Österholm P (2012) The limited usefulness of macroeconomic Bayesian vars when estimating the probability of a US recession. J Macroecon 34(1):76–86
86. Hamilton JD, Lin G (1996) Stock market volatility and the business cycle. J Appl Econom 11(5):573–593
87. Estrella A, Mishkin FS (1998) Predicting US recessions: financial variables as leading indicators. Rev Econ Stat 80(1):45–61
88. Qi M (2001) Predicting US recessions with leading indicators via neural network models. Int J Forecast 17(3):383–401
89. Berge TJ (2015) Predicting recessions with leading indicators: model averaging and selection over the business cycle. J Forecast 34(6):455–471
90. O'hagan A, Leonard T (1976) Bayes estimation subject to uncertainty about parameter constraints. Biometrika 63(1):201–203
91. Cramer JS (1987) Mean and variance of r2 in small and moderate samples. J Econom 35(2–3):253–266
92. Carrodus ML, Giles DE (1992) The exact distribution of r2 when the regression disturbances are autocorrelated. Econ Lett 38(4):375–380
93. Kleinberg J (2003) Bursty and hierarchical structure in streams. Data Min Knowl Discov 7(4):373–397
94. Leskovec J, Backstrom L, Kleinberg J (2009) Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 497–506
95. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022
96. Dou W, Wang X, Chang R, Ribarsky W (2011) Paralleltopics: a probabilistic approach to exploring document collections. In: 2011 IEEE conference on visual analytics science and technology (VAST). IEEE Press, New York, pp 231–240
97. Serrano MÁ, Boguná M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci 106(16):6483–6488
98. Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70(6):066111

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com