



OPEN

Fluctuations and extreme events in the public attention on Italian legislative elections

Andrea Auconi¹, Lorenzo Federico^{2,3✉}, Gianni Riotta² & Guido Caldarelli^{1,4,5,6}

The share of social media attention to political candidates was shown to be a good predictor of election outcomes in several studies. This attention to individual candidates fluctuates due to incoming daily news and sometimes reflects long-term trends. By analyzing Twitter data in the 2013 and 2022 election campaign we observe that, on short timescales, the dynamics can be effectively characterized by a mean-reverting diffusion process on a logarithmic scale. This implies that the response to news and the exchange of opinions on Twitter lead to attention fluctuations spanning orders of magnitudes. However, these fluctuations remain centered around certain average levels of popularity, which change slowly in contrast to the rapid daily and hourly variations driven by Twitter trends and news. In particular, on our 2013 data we are able to estimate the dominant timescale of fluctuations at around three hours. Finally, by considering the extreme data points in the tail of the attention variation distribution, we could identify critical events in the electoral campaign period and extract useful information from the flow of data.

The recent technological revolution made it possible to store unprecedented information regarding society. Data from all kinds of media, including e-mails, phone calls, online purchases, travel info and more, are stored, forming what has been called the data deluge¹. This availability of data has offered more opportunities for advanced mathematical techniques to be applied in the modeling of social and political phenomena, towards a quantitative description of human behaviour and its multi-layer interactions within society^{2–4}. The classical definition of complexity as “more is different”⁵ manifests here in the fact that social interactions, which typically involve just two or only a few individuals, can propagate their effect through a network to generate collective behaviour in a way that is reminiscent of phase transitions in statistical physics^{6,7}.

Methods of statistical physics and network theory^{8,9} have already found application in social sciences for the study of information spreading and fake news^{10–12}, and in particular to characterise the relation between news diffusion properties and the underlying network structure. The latter is usually inferred from interactions on platforms like Twitter and Facebook. On Twitter, the clustering around influential people is particularly strong¹³, as well as the emergence of closed “echo chambers” of people mutually reinforcing their opinions^{14,15}, so that a coarse-grained description in terms of communities and narratives was possible. In particular, studies based on Twitter data elucidated contagion mechanisms in the Arab Spring demonstrations¹⁶, Occupy Wall Street¹⁷, and the Spanish “Indignados”¹⁸. Also, many tried to extract information from social networks to get some anticipation on the outcome of online electoral campaigns in various countries including USA^{19–25}, Australia^{26,27}, Norway²⁸, Spain²⁹, Italy^{30,31}, France³², and UK³³.

In this paper, we study the fluctuation properties of the attention towards political parties on Twitter in the weeks preceding the Italian political elections of 2013 and 2022. In particular our question is: how can we improve our models of stochastic varying attention online, so that we can better connect this to election outcomes and possibly other phenomena? We find that, on short timescales of minutes up to a few days, the dynamics is well described by a mean-reverting diffusion process on a logarithmic scale, meaning that the reaction to news and exchange of opinions on Twitter produces fluctuations of attention varying over orders of magnitudes but still anchored around some set of average levels of popularity which change only slowly compared to the daily and hourly variations due to news and trends on Twitter. We find that this minimal description employing just two parameters, one for the timescale of mean-reversion and another for the amplitude of fluctuations, is sufficient

¹Department of Molecular Sciences and Nanosystems, Ca' Foscari University of Venice, 30172 Venice, Italy. ²LUISS Data Lab, Viale Pola 12, 00198 Rome, Italy. ³Department of Political Sciences, LUISS University, Viale Romania 32, 00197 Rome, Italy. ⁴European Centre for Living Technology, 30124 Venice, Italy. ⁵Institute for Complex Systems, Consiglio Nazionale delle Ricerche, UoS Sapienza, 00185 Rome, Italy. ⁶London Institute for Mathematical Sciences, Royal Institution, London W1K2XF, UK. ✉email: lfederico@luiss.it

to interpolate the average squared logarithmic displacement of the share of Twitter attention, which is a basic measure of uncertainty on the future level of attention. By calibrating this model to our time series data we find that the dominant timescale of fluctuations can be estimated at around three hours.

Regarding the distribution of these squared log displacements, however, we observe some larger tails than the Gaussian model would typically show. Indeed, like in many other social phenomena, we also find some extreme events, here as short periods of collective attention³⁴. These are relatively rare and happen in a small absolute number in our data. Therefore, further modeling in terms of a jump–diffusion process and a fat-tailed distribution is not statistically feasible. Still, we investigate these individual jumps in attention and find that they correspond to significant unexpected events in the political campaign. Our analysis of fluctuations on short timescales of minutes up to a few days complements the studies of trends and inertia in the attention dynamics over longer timescales^{31,35–41}.

Methods

The datasets

The first and main dataset used here consists of around 3.5 M tweets recorded from 1 January 2013 to 22 February 2013 before the general Italian political elections which took place on the 24 and 25 February 2013. This dataset was created and already used for a different analysis in³¹. Tweets were selected using the Twitter free API in *stream* mode with Italian language as only filter, and then classified based on if they mention political leaders' family names or their political party with or without hashtag. The free API provides only a subset of the whole tweets population, and this determines an increased estimation noise especially relevant for small filtering timescales as discussed below. Similarly the second dataset for the general Italian political elections of 25 September 2022 consists of around 2.5 M Tweets recorded from 6 September 2022 to 23 September 2022 selected according to a list of hashtags provided in⁴², and then classified based on if they mention the leaders and the six main political parties, such keywords provided in the Supplementary file “Keywords”.

The share of Twitter attention

Our data consists of a time series of tweets regarding politicians in Italy during the weeks preceding the political elections of 2013 and 2022. For a given political candidate i the time series is a list of events' times $[t_1, t_2, \dots, t_{n_i}]$ corresponding to tweets mentioning that particular candidate or its political party. We express this equivalently with a measurement process $m(t)$ making unit jumps for each event, namely $dm(t) = 1$ if t is in the list of events' times and $dm(t) = 0$ otherwise. The differentials $dm(t) = 1$ can be interpreted as the limit of a time discretization of vanishing size. We wish to estimate an instantaneous twitting rate $y(t)$ from the measurement process $m(t)$, by using known results in the field of stochastic dynamics. Indeed, this problem is generally formulated in the language of stochastic filtering^{43,44}, where a likelihood function is dynamically updated with respect to a model for the stochastic dynamics of the unobserved process $y(t)$ and its measurement probability $p(dm|y)$. However, we wish not to make strong modeling assumptions at this stage, therefore we considered a linear filter of the kind

$$y(t) = \beta \int_{-\infty}^t dm(t') e^{-\beta(t-t')}, \quad (1)$$

with a timescale β^{-1} which should be large enough to contain a sufficiently large number of events (for a precise rate estimation), but also sufficiently small compared to the natural time of tweet rate variations. As the latter is not yet determined, we practically fix $\beta^{-1} = 20$ minutes and later vary it to check that results do not significantly depend on it. Similarly, the results should be invariant if taking a time binning approach with size β^{-1} instead of the corresponding linear filter. Let us denote the expectation value of a quantity as $\mathbb{E}[\dots]$. We can express Eq. (1) equivalently in differential form as $dy = \beta(dm - y dt)$, and we see that if the twitting rate is constant, i.e. $\mathbb{E}[dm] = r dt$, then at steady state ($\mathbb{E}[dy] = 0$) we have $\mathbb{E}[y] = r$ as expected. The fluctuations around this average level due to the discreteness of events is calculated as $\text{Var}[y] = r\beta/2$.

We are here interested in the share of Twitter attention on the various political candidates, as this may indicate voting intentions in the population. Therefore we normalize each y_i by its total among the candidates by defining $x_i = y_i / (\sum_j y_j)$. This definition of x_i as a share of Twitter attention also works as a de-trending mechanism. The twitting rate has a prominent daylight dependence since the activity is much reduced at night. While some noise due to the lower rates in the night affects x , we nevertheless check the stability of our results by removing some of the night hours from the y integral. Finally, let us note that considering the twitting rate in relative terms has the consequence of showing apparent drops due to spikes in other processes.

Geometric Ornstein–Uhlenbeck process

Consider the following stochastic differential equation in the Itô interpretation⁴⁵,

$$dx = \sigma x dW - \alpha x \ln \left(\frac{x}{\mu} \right) dt, \quad (2)$$

where dW denote Brownian motion increments satisfying $\mathbb{E}[dW(t)] = 0$, $\mathbb{E}[dW(t)dW(t')] = \delta_{tt'} dt$, and the Dirac delta is here $\int_0^T \delta_{tt'} = \mathbb{I}_{t \in [0, T]}$ in terms of the indicator function $\mathbb{I}_{t \in [0, T]}$. Using Itô's Lemma⁴⁵ we find

$$d \ln x = \sigma dW - \alpha \ln \left(\frac{x}{\mu} \right) dt - \frac{\sigma^2}{2} dt, \quad (3)$$

that can be rewritten in terms of $\tilde{x} \equiv \ln x$ as

$$d\tilde{x} = \sigma dW - \alpha(\tilde{x} - \tilde{\mu})dt, \tag{4}$$

where $\tilde{\mu} = \ln \mu - \frac{\sigma^2}{2\alpha}$. We see from the definition of \tilde{x} and its dynamics in Eq. (4) that x follows an Ornstein–Uhlenbeck process⁴⁶ in logarithmic scale, which we here call Geometric Ornstein–Uhlenbeck (GOU) process. The GOU process combines the logarithmic scale diffusion of the geometric Brownian motion with the linear mean-reverting property of the Ornstein–Uhlenbeck process. These two features are governed by the volatility parameter σ and the inverse correlation time α . The Ornstein–Uhlenbeck process admits a stationary Gaussian density $p(\tilde{x}) = \mathcal{N}(\tilde{\mu}, \frac{\sigma^2}{2\alpha})$ with mean $\mathbb{E}[\tilde{x}] = \tilde{\mu}$ and variance $\mathbb{E}[\tilde{x}^2] - (\mathbb{E}\tilde{x})^2 = \frac{\sigma^2}{2\alpha}$. The average level of the x process is given by the Gaussian integral $\mathbb{E}[x] = \int d\tilde{x} p(\tilde{x}) e^{\tilde{x}} = \mu e^{-\frac{\sigma^2}{4\alpha}}$.

The squared log-displacement

Let us consider the squared log-displacement variable

$$s(\tau) = \left[\ln \left(\frac{x(\tau)}{x(0)} \right) \right]^2, \tag{5}$$

where for simplicity we drop the explicit dependence on t and imagine to sample the process at $t = 0$ from the stationary density. From the above definitions we can write $s(\tau) = (\tilde{x}(\tau) - \tilde{x}(0))^2$, consider the well-known solution to the Ornstein–Uhlenbeck process,

$$\tilde{x}(\tau) = (\tilde{x}(0) - \tilde{\mu})e^{-\alpha\tau} + \tilde{\mu} + \sigma \int_0^\tau e^{-\alpha(\tau-t)} dW(t), \tag{6}$$

and use it to evaluate Eq. (5). Then noting that $\mathbb{E}[(\tilde{x}(0) - \tilde{\mu})^2] = \frac{\sigma^2}{2\alpha}$ in the stationary density, and computing the expectations as usual from the Brownian motion covariance, we obtain

$$\mathbb{E}[s(\tau)] = \frac{\sigma^2}{\alpha}(1 - e^{-\alpha\tau}), \tag{7}$$

which characterizes the scaling of the mean square log-displacement with the time interval τ and its exponential saturation due to the linear mean-reversion.

Fitting the data

The quantity $x(t)$ that we choose to analyze the Twitter share of attention dynamics is the empirical average square log displacement. Accordingly, the model to interpret its dependence on the time interval is the GOU process described above. In particular, we calibrate the two parameters σ and α in the theoretical curve of Eq. (7), $\mathbb{E}[s(\tau_j)] \equiv \mathbb{E}[s(\tau_j)]_{\sigma,\alpha}$, from samples of the empirical average square log-displacement $\tilde{\mathbb{E}}[s(\tau_j)]$ for a set of time intervals, $\tau_j = \tau_0 \lfloor 1.2^j \rfloor$, with the shortest $\tau_0 = 30$ min taken large enough so that the impact of the discreteness of events is small, and up to $\tau_j < 30$ hours as to be still much smaller than the total observation period. We also require $\tau_0 > \beta^{-1}$ to reduce the overlapping of the exponential kernels in the tweet rate estimation. The impact of the discreteness can then be taken into account approximately (by neglecting normalization correlations) to first order in the empirical average rate $\tilde{\mathbb{E}}[y_i]$ and variance $\tilde{\text{Var}}[y_i]$ by subtracting to the empirical curve $\tilde{\mathbb{E}}[s(\tau_j)] \rightarrow \tilde{\mathbb{E}}[s(\tau_j)] - \tilde{\mathbb{E}}[s]^{(b)}$ a background noise

$$\tilde{\mathbb{E}}[s]^{(b)} = \frac{\beta}{\tilde{\mathbb{E}}[y_i]} \left(1 + \frac{\tilde{\text{Var}}[y_i]}{(\tilde{\mathbb{E}}[y_i])^2} \right), \tag{8}$$

where the Log-Normal distribution for y was assumed.

The fit is then performed by minimizing the mean squared error, $\min_{\sigma,\alpha} \sum_j \left(\tilde{\mathbb{E}}[s(\tau_j)] - \mathbb{E}[s(\tau_j)]_{\sigma,\alpha} \right)^2$.

Once the parameters $\tilde{\sigma}_i$ and $\tilde{\alpha}_i$ are fitted for each political candidate i , we study if the discrepancy from the fitted curve has some common structure, which would signal for improper modeling assumptions. This is done by normalizing the empirical expectation $\tilde{\mathbb{E}}[s(\tau_j)]_i$ for each candidate i by the fitted $\frac{\tilde{\sigma}_i^2}{\tilde{\alpha}_i}$, and then rescaling the time intervals differently for each candidate by the fitted $\tilde{\alpha}_i$, namely

$$z_i(\tilde{\alpha}_i \tau_j) = \tilde{\mathbb{E}}[s(\tau_j)]_i \frac{\tilde{\alpha}_i}{\tilde{\sigma}_i^2}, \tag{9}$$

where τ_j are the points in the original time intervals discretization. With this we obtain normalized samples $[\tau_k, z(\tau_k)]_i$ which are now comparable between political candidates.

Alternative model with jumps

The exponential saturation of the mean square log displacement does not uniquely identify the GOU process. Indeed consider a non-diffusive process that jumps between two discrete states $x \in \{1, c\}$ with symmetric rate r . Its conditional probability $p \equiv p(x(\tau) = 1 | x(0) = 1)$ evolves with $\partial_t p = r(1 - 2p)$ which gives $p = (1 + e^{-2r\tau})/2$ from the initial condition $p = 1$. We immediately find $\mathbb{E}[s(\tau)] = \frac{(\ln c)^2}{2}(1 - e^{-2r\tau})$, which means that the exponential saturation is obtained also by a qualitatively different process not based on Brownian motion, and potentially many more examples are possible. There is therefore a degeneracy in the models that fit the exponential curve of Eq. (7). This degeneracy can be resolved by the distribution of the squared log increments, as for example, this is smooth in the GOU case while it is a single peak in the jump process described above. In this simple jump process and also in more general processes combining diffusion with jumps of uncertain size, the distribution of squared log increments will be characterized by a fat tail for small intervals τ because jumps are statistically incompatible with continuous diffusion.

Anomalous diffusion

As mentioned above, the linear scaling of the squared log displacement for small time intervals τ is consistent with multiple types of diffusion and jump processes. One may ask what model processes instead do not follow this linear scaling law, and the answer is in the theory of anomalous diffusion^{47–49}. We argue that this latter more complicated framework is not necessary for our data as we show that, if the twitting rate is filtered with a large enough timescale β^{-1} and for a sizeable range of this parameter, the process is well described by a standard diffusion like the GOU process.

Checking the distribution

After the parameters $\tilde{\sigma}_i$ and $\tilde{\alpha}_i$ are fitted on the empirical average square log-displacement, we can dig a bit deeper and check if the empirical probability distribution of the square log displacements $s_i(\tau_j)$ is comparable with that of the GOU, which is a chi-squared distribution being the square of a Gaussian. For looking at this on a comparable scale between political candidates we consider the normalized increments

$$\tilde{z}_i(\tau_j) = \frac{s_i(\tau_j)}{\mathbb{E}[s(\tau_j)]|_{\tilde{\sigma}_i, \tilde{\alpha}_i}}. \quad (10)$$

Further detrending

We wish to use as few parameters as possible to keep the description simple, so we do not add additional detrending on the share of attention $x(t)$ to realize the main text figures. For the sake of completeness, however, we observe that for some political candidates, and especially for the case of the 2022 elections, where our observation period is shorter, a non-negligible drift in popularity is found over the whole period under analysis. This could confound the fit to the mean-reverting process, as the long-range variation is interpreted as a longer correlation time α^{-1} . For this, we introduce a simple detrending of the form $a_i e^{b_i(t-t_0)}$ which improves the fit to the GOU as is shown in the supplementary figures. For our data this period-long drift results to be small enough that the GOU fit to the square log-displacement is good even without detrending when the time intervals considered are between a few minutes and up to a few days. This limit is taken also in order to have enough samples for estimating the parameters α and σ , and accordingly the calibration intervals τ should be much shorter compared to the full time span of the data. Furthermore, note that the curve calibration does not involve the average $\tilde{\mu}$ as seen in Eq. (7). Therefore it seems reasonable when performing predictions to estimate $\tilde{\mu}$ from an interval of around the calibration timescale rather than the whole history.

Results

Logarithmic fluctuations and mean reversion

The share of attention to Italian political parties on Twitter has already been shown to be a good indicator of election outcomes^{30,31}. We here discuss how a description in terms of the simple GOU process, which is a model of mean-reverting diffusion in logarithmic scale discussed in the Methods section, fits the short-term empirical fluctuations in the share of attention derived from our data for the 2013 and 2022 Italian political elections.

The fit of the mean square log displacement to the GOU theoretical curve gives the parameters in table 1 for the four main parties in 2013 for which the statistics is sufficient. $\tilde{\alpha}$ is the mean-reverting parameter and $\tilde{\sigma}$ is the volatility parameter controlling the amplitude of fluctuations. We see that the timescale of mean-reverting

Political party	$\tilde{\alpha}$	$\tilde{\sigma}$
PD	0.301	0.282
SC	0.298	0.235
PdL	0.255	0.239
M5S	0.297	0.368

Table 1. Parameter fit for the GOU process in the 2013 Italian political elections, with a filtering timescale of $\beta^{-1} = 20$ min.

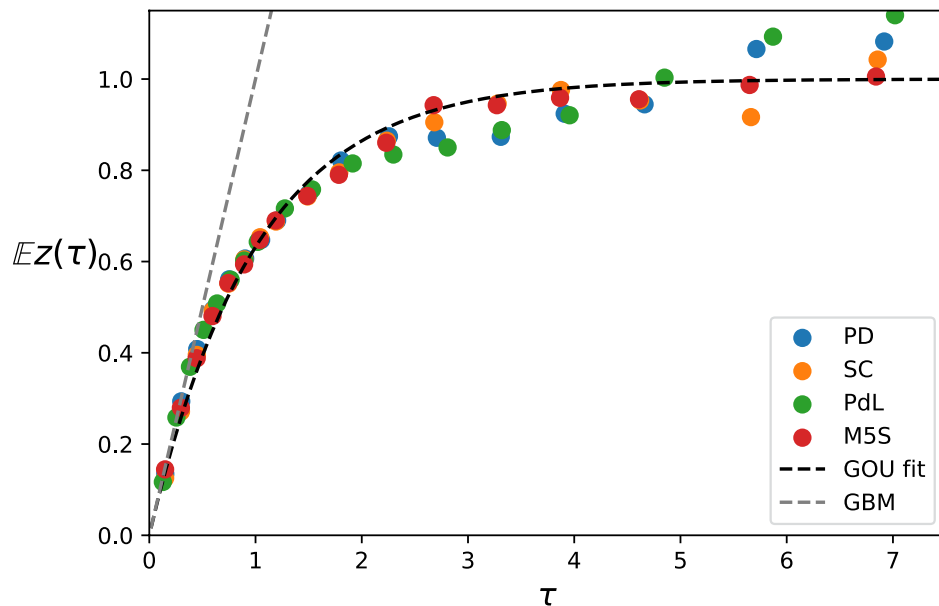


Figure 1. Fit to the GOU process Empirical mean square log-displacement in normalized units for the four main political parties in the Italian 2013 elections, and the corresponding theoretical curves. The filtering timescale used for the instantaneous twitting rate is $\beta^{-1} = 20$ min. We see that the GOU model, a mean-reverting diffusion process in logarithmic scale with just two parameters is able to fit the empirical curves on short timescales. GBM is Geometric Brownian motion. Precise definitions in the Methods section.

is around $\tilde{\alpha}^{-1} \approx 3$ hours for all parties, which gives an idea of the typical reaction and debate time on Twitter to incoming news. This mean-reverting timescale estimate of roughly three hours results to be stable under changes in the filtering timescale β^{-1} , see supplementary Fig. A1. The volatility results to be larger for the M5S five-star movement party, and this can partly be explained by their twitting rate being smaller and absolute variations being amplified in logarithmic scale.

By normalizing the twitting rate and rescaling time according to the obtained GOU fit, we visualize the structure of the fit in Fig. 1. Some slight deviation from the theoretical curve suggests the presence of nonlinearities or some drift structure visible already at the calibration timescale. Actually, it could seem like a rather strong assumption for the twitting rate to fluctuate around an underlying average which is constant during the few weeks preceding the elections, as a political campaign typically intensifies when approaching election dates exactly to shift the political attention towards their candidates. Indeed our calibration is limited to timescales between a few minutes up to a few days so that trends on longer timescales are not considered. For our particular dataset, however, we find that removing the trend with a simple exponential fit does not substantially improve the GOU curve fit of Fig. 1, see Supplementary Fig. A2, meaning that the drift structure is more complex than our single-timescale description. Further characterizing these effects is not statistically feasible with our small dataset, and we consider the minimal description in terms of the GOU process appropriate and the fit in Fig. 1 satisfactory.

Regarding the distribution of the squared log-displacement, the GOU process would prescribe it to be a chi-squared distribution. In Fig. 2 we see some deviation from the theoretical curve at the large values tail, which cannot be explained from the diffusive statistics and are more properly described as jumps. Indeed, as discussed in the Methods section, also a jump process can have a mean square log-displacement analogous to the GOU process, and the degeneracy is exactly removed by looking at the distribution. However, these jumps happen in a relatively small number and it is not feasible to add more parameters to model a jump distribution on top of the diffusion from this small dataset, and we therefore consider these just as outliers or extreme events.

Let us also mention that we observe some skewness in the log-displacement distribution which is understood as typically the reaction to events is faster than the relaxation from the excited states.

Extreme events

Let us here consider some of these large twitting rate jump events, as they reflect the community's reaction to the most impactful news.

The first one we list happened on the 12/02/2013 and affected the PdL party of Silvio Berlusconi. In particular, a satirical comedian well-known in Italy went to the Sanremo music festival for a comedy segment which particularly targeted Berlusconi. Some in the audience started shouting and interrupted the comedian, which sparked a broader discussion in the public opinion and likely also on Twitter about free speech and the appropriateness of political satire in the national music festival.

Another one regards the PD democratic party when on the 07/01/2013 the candidate president Pierluigi Bersani was hosted in the popular television program "Otto e Mezzo" to discuss his political agenda. The following day Berlusconi participated in the same tv program and a comparable spike happened in the share of attention.

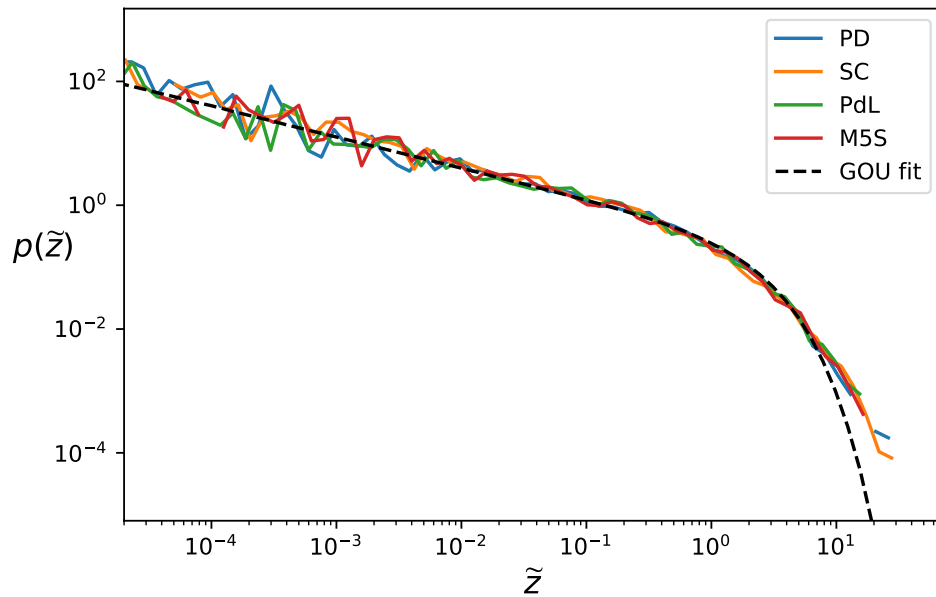


Figure 2. Comparing the distribution. Empirical distribution of the squared log-displacement in normalized units and comparison to the GOU model fit from Fig. 1. The fat tails are explained by few extreme events. Python “powerlaw”⁵⁰ package was used for plotting. The time interval is $\tau = 180$ min, larger than the filtering timescale and comparable with the mean-reversion timescale.

From the definition we see that the events which impact most the square log-displacement on the shortest timescale β^{-1} are the unexpected news. As an example consider the very popular tv debate between Berlusconi and its most critical journalist Marco Travaglio, and while it provoked a massive debate and interest on Twitter, the program was relatively long and it was planned weeks ahead, so that the Twitter reactions spread over a longer period of time so that twitting rate variations were relatively moderate.

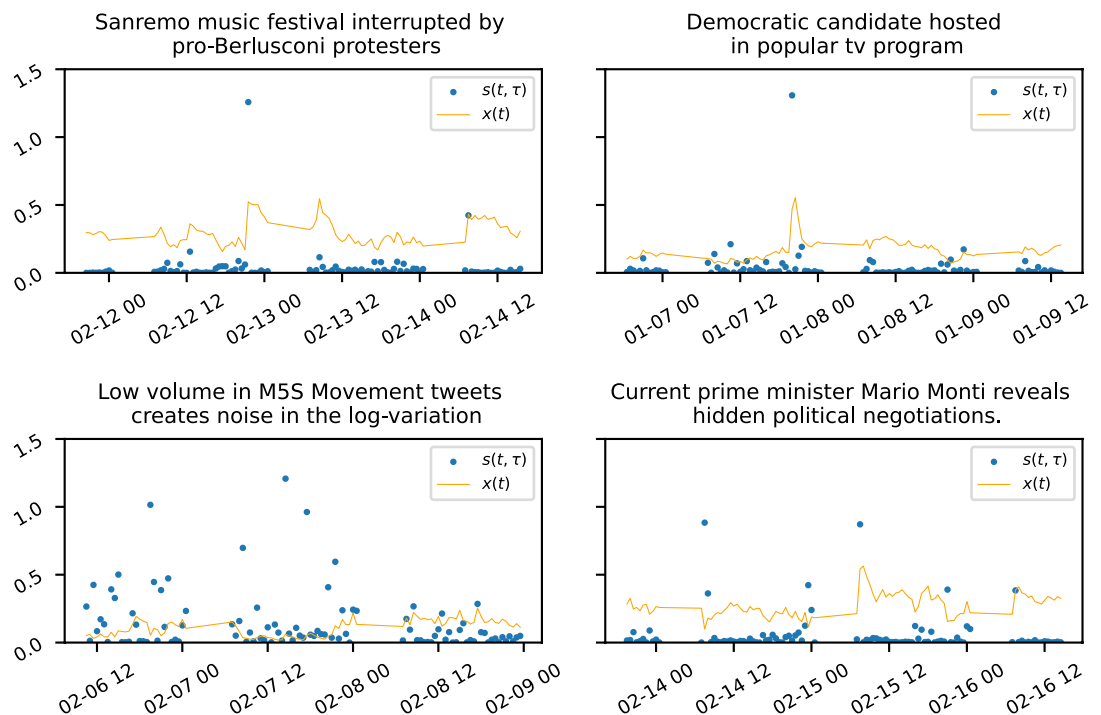


Figure 3. Examples of extreme events. $x(t)$ is the share of attention at time t , and $s(t, \tau)$ is the corresponding squared log-displacement, here for an interval $\tau = 30$ min and with filtering timescale $\beta^{-1} = 20$ min. We see how significant and unexpected events in the electoral campaign are reflected in spikes of $x(t)$.

Another event we discuss regards the SC party, that from its official Twitter account in the morning of 15/02/2013 revealed that their candidate Mario Monti had been secretly offered important political roles from other parties in exchange for him renouncing to run for president. Understandably, this sparked quite an immediate reaction.

We note that for the anti-establishment five star movement, which actually was the surprise of the elections results, Twitter was not the main platform for propaganda and discussion as they did use the movement's website. As a consequence, the twitting rate was often so low that small absolute variations can have a big impact in the logarithmic variations, as is shown in Fig. 3 together with the other events discussed here.

What changed in 2022

We have a much shorter dataset for the 2022 elections, and only for the two weeks preceding the elections which were strongly characterized by two events, namely the television debate between the two main candidates Enrico Letta and Giorgia Meloni, and the strong personal accusations between the other two candidates Matteo Renzi and Giuseppe Conte. These two events concentrated much of the public attention in our data timeframe, and as a consequence we observe a lack of saturation and the fitted timescale of mean-reverting is larger (around two times) compared to the 2013 elections. The fit to the GOU process is still satisfactory but less accurate, see Supplementary Fig. A3.

Out-of-sample predictions

We study the dynamics of attention on short timescales and in this limit we identify fluctuations and mean-reversion as the two main dynamical features. The GOU model can then be used for predicting the attention dynamics on timescales comparable to the mean-reverting timescale α , which was here estimated to be of around 3 h. As an example, we consider the last day before the elections 22 February 2013 (23 Feb was pre-election silence) at four different time instants, and visualize the expectation and standard deviation obtained from the GOU model. The parameters are obtained as above but only from data up to the prediction time instant, and with an instantaneous log-mean $\tilde{\mu}$ for mean-reversion estimated from the previous 40 hours which is the largest calibration interval. The results are shown in Fig. 4. The well-known expectations of the Ornstein–Uhlenbeck process⁴⁶ are such that, given an initial state $\tilde{x}(0)$, the expected state after a time interval τ is $\mathbb{E}[\tilde{x}(\tau)|\tilde{x}(0)] = (\tilde{x}(0) - \tilde{\mu})e^{-\alpha\tau} + \tilde{\mu}$, and its variance is $\text{Var}[\tilde{x}(\tau)|\tilde{x}(0)] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\tau})$.

Discussion

In this paper we present an analysis on two distinct dataset of Twitter messages time series corresponding to different Italian political elections. Our study reveals that, on short timescales from a few minutes to a few days, the dynamics of political discussions on Twitter exhibit characteristics consistent with a Geometric Ornstein–Uhlenbeck (GOU) evolution, that is a linear mean-reverting diffusion process on a logarithmic scale. Crucially, this

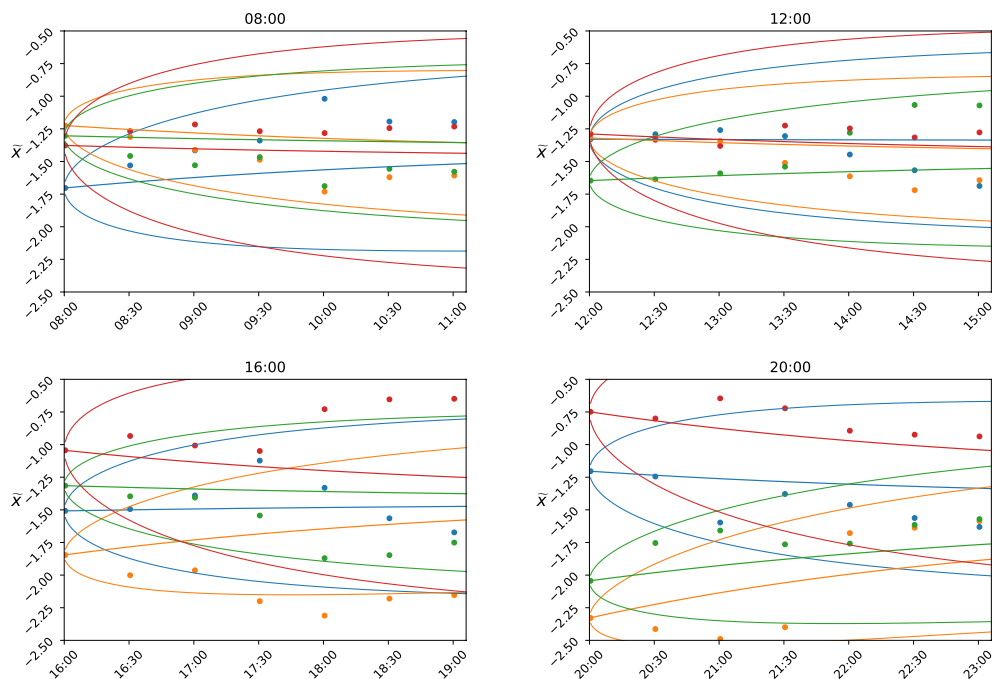


Figure 4. Predictions using the GOU model. Time series predictions based on the GOU model with the parameters estimated immediately before prediction, at four different times during the day of 22 February 2013. The scatter plot is the realized \tilde{x} , while the three solid lines are the expectation and ± 2 standard deviations. The colors are as above: blue is 'PD', orange is 'SC', green is 'PdL', red is 'M5S'.

minimal description involves a single timescale which we estimate from our data to be of around three hours. Small discrepancies from the fitted curve reflect the non-modeled multiscale properties of the dynamics. This unique modeling choice enables us to capture the nuanced patterns of attention and opinion exchange in the ever-changing landscape of political conversations. The approach presented here contributes to the problem of modeling short-term fluctuations in Twitter attention by estimating a timescale of mean-reversion which enables the detection of changes in the location of the longer-term process net of short-lived fluctuations; how to model election outcomes as a function of this Twitter attention relies instead on other studies^{35,51}. We have here been dealing with a politics-only dataset, therefore we cannot establish whether the GOU model is peculiar of the political attention or rather a more general form of attention dynamics.

One notable aspect of our analysis is the identification of extreme events within the political Twitter data. These events, characterized by significant deviations from the linear mean-reverting behavior, underscore the non-linear and unpredictable nature of political discussions on the platform. Understanding and quantifying these extreme events are crucial for a comprehensive analysis of the evolving political discourse. The application of our findings in election forecasting holds substantial promise. By incorporating the GOU evolution model, we gain a more nuanced understanding of how political sentiments and attention levels fluctuate over time. This insight can enhance the accuracy of election predictions by accounting for the inherent volatility in public opinion, especially during critical periods such as debates, policy announcements, or unexpected events. Indeed, the GOU model allows for real-time monitoring of Twitter data, providing a dynamic and adaptive framework for assessing the ebb and flow of political discussions. This real-time capability enables us to identify emerging trends and potential turning points in public opinion, offering valuable insights for election forecasting models that seek to capture the zeitgeist of the electorate. Our approach complements the studies of the attention to politicians on longer timescales where fluctuations become negligible and the dynamics is described by trends and inertial effects^{31,35–41}. While our study presents a novel framework for understanding political Twitter dynamics, there are challenges and avenues for further exploration. Future research could delve into refining the GOU model parameters and especially determining if the three hours timescale identified here changes in other datasets and situations, and more in general investigating the transferability of the model across different political landscapes.

Data availability

The 2013 Data are available from Figshare as published in Ref.³¹, <http://dx.doi.org/10.6084/m9.figshare.1437740>. The 2022 dataset is available from the corresponding author on reasonable request. The python code developed for the data analysis in this manuscript is publicly available on https://github.com/AndreaAuconi/Twitter_time_series_analysis.

Received: 4 January 2024; Accepted: 4 August 2024

Published online: 01 October 2024

References

- Bell, G., Hey, T. & Szalay, A. Beyond the data deluge. *Science* **323**, 1297–1298 (2009).
- Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
- De Domenico, M., Granell, C., Porter, M. A. & Arenas, A. The physics of spreading processes in multilayer networks. *Nat. Phys.* **12**, 901–906 (2016).
- Caldarelli, G., Wolf, S. & Moreno, Y. Physics of humans, physics for society. *Nat. Phys.* **14**, 870–870 (2018).
- Anderson, P. W. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* **177**, 393–396 (1972).
- Bialek, W. *et al.* Social interactions dominate speed control in poising natural flocks near criticality. *Proc. Natl. Acad. Sci.* **111**, 7212–7217 (2014).
- Levy, M. Stock market crashes as social phase transitions. *J. Econ. Dyn. Control* **32**, 137–155 (2008).
- Caldarelli, G. *Scale-free Networks: Complex Webs in Nature and Technology* (Oxford University Press, 2007).
- Barabási, A.-L. Taming complexity. *Nat. Phys.* **1**, 68–70 (2005).
- Zurek, W. H. *Complexity, Entropy and the Physics of Information* (CRC Press, 2018).
- Zhang, Z.-K. *et al.* Dynamics of information diffusion and its applications on complex networks. *Phys. Rep.* **651**, 1–34 (2016).
- Lazer, D. M. *et al.* The science of fake news. *Science* **359**, 1094–1096 (2018).
- Sacco, P. L., Gallotti, R., Pilati, F., Castaldo, N. & De Domenico, M. Emergence of knowledge communities and information centralization during the covid-19 pandemic. *Soc. Sci. Med.* **285**, 114215 (2021).
- Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
- Del Vicario, M. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **113**, 554–559 (2016).
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A. & Moreno, Y. The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 1–7 (2011).
- Conover, M. *et al.* Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media* vol. 5, 89–96 (2011).
- González-Bailón, S., Borge-Holthoefer, J. & Moreno, Y. Broadcasters and hidden influentials in online protest diffusion. *Am. Behav. Sci.* **57**, 943–965 (2013).
- Adamic, L. A. & Glance, N. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43 (2005).
- Diakopoulos, N. A. & Shamma, D. A. Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI Conference On Human Factors in Computing Systems* 1195–1198 (2010).
- DiGrazia, J., McKelvey, K., Bollen, J. & Rojas, F. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLoS ONE* **8**, e79449 (2013).
- Bekafigo, M. A. & McBride, A. Who tweets about politics? political participation of twitter users during the 2011 gubernatorial elections. *Soc. Sci. Comput. Rev.* **31**, 625–643 (2013).
- Badawy, A., Ferrara, E. & Lerman, K. Analyzing the digital traces of political manipulation: The 2016 Russian interference twitter campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* 258–265 (IEEE, 2018).

24. Bovet, A. & Makse, H. A. Influence of fake news in twitter during the 2016 us presidential election. *Nat. Commun.* **10**, 7 (2019).
25. Zhou, Z., Serafino, M., Cohan, L., Caldarelli, G. & Makse, H. A. Why polls fail to predict elections. *J. Big Data* **8**, 1–28 (2021).
26. Gibson, R. K. & McAllister, I. Does cyber-campaigning win votes? Online communication in the 2004 Australian election. *J. Elect. Public Opin. Parties* **16**, 243–263 (2006).
27. Bruns, A. & Stieglitz, S. Quantitative approaches to comparing communication patterns on twitter. *J. Technol. Hum. Serv.* **30**, 160–185 (2012).
28. Enli, G. S. & Skogerbø, E. Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. *Inform. Commun. Soc.* **16**, 757–774 (2013).
29. Borondo, J., Morales, A. J., Losada, J.-C. & Benito, R. M. Characterizing and modeling an electoral campaign in the context of twitter: 2011 Spanish presidential election as a case study. *Chaos Interdiscip. J. Nonlinear Sci.* **22**, 023138 (2012).
30. Caldarelli, G. *et al.* A multi-level geographical study of Italian political elections from twitter data. *PLoS ONE* **9**, e95809 (2014).
31. Eom, Y.-H., Puliga, M., Smailović, J., Mozetič, I. & Caldarelli, G. Twitter-based analysis of the dynamics of collective attention to political parties. *PLoS ONE* **10**, e0131184 (2015).
32. Ferrara, E. Disinformation and social bot operations in the run up to the 2017 french presidential election. Tech. Rep. 2995809, SSRN Scholarly Paper (2017).
33. Llewellyn, C. & Cram, L. Brexit? analyzing opinion on the uk-eu referendum within twitter. In *Proceedings of the International AAAI Conference on Web and Social Media* vol. 10, 760–761 (2016).
34. De Domenico, M. & Altmann, E. G. Unraveling the origin of social bursts in collective attention. *Sci. Rep.* **10**, 4629 (2020).
35. Beauchamp, N. Predicting and interpolating state-level polls using twitter textual data. *Am. J. Polit. Sci.* **61**, 490–503 (2017).
36. Aletti, G., Crimaldi, I. & Saracco, F. A model for the twitter sentiment curve. *PLoS ONE* **16**, e0249634 (2021).
37. Wang, H., Can, D., Kazemzadeh, A., Bar, F. & Narayanan, S. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations* 115–120 (2012).
38. Klačnja, M., Barberá, P., Beauchamp, N., Nagler, J. & Tucker, J. A. Measuring public opinion with social media data (2015).
39. Shugars, S. & Beauchamp, N. Why keep arguing? predicting engagement in political conversations online. *SAGE Open* **9**, 2158244019828850 (2019).
40. Beck, N. Comparing dynamic specifications: The case of presidential approval. *Polit. Anal.* **3**, 51–87 (1991).
41. Gelman, A. & King, G. Why are American presidential election campaign polls so variable when votes are so predictable?. *Br. J. Political Sci.* **23**, 409–451 (1993).
42. Federico, L., Mounim, A. & Caldarelli, G. E. A. Multi-scale analysis of the community structure of the twitter discourse around the Italian general elections of september 2022. *Sci. Rep.* <https://doi.org/10.1038/s41598-024-65564-6> (2024).
43. Kallianpur, G. *Stochastic Filtering Theory* Vol. 13 (Springer, 2013).
44. Petris, G., Petrone, S. & Campagnoli, P. *Dynamic Linear Models with R* (Springer, 2009).
45. Karatzas, I. & Shreve, S. *Brownian Motion and Stochastic Calculus* Vol. 113 (Springer, 2012).
46. Risken, H. & Risken, H. *Fokker-Planck Equation* (Springer, 1996).
47. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: A fractional dynamics approach. *Phys. Rep.* **339**, 1–77 (2000).
48. Chechkin, A. V., Metzler, R., Klafter, J. & Gonchar, V. Y. Introduction to the theory of lévy flights. In *Anomalous transport: Foundations and applications* (eds Klages, R. *et al.*) 129–162 (Wiley, 2008).
49. Klages, R., Radons, G. & Sokolov, I. M. *Anomalous Transport* (Wiley Online Library, 2008).
50. Alstott, J., Bullmore, E. & Plenz, D. powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE* **9**, e85777 (2014).
51. Liu, R., Yao, X., Guo, C. & Wei, X. Can we forecast presidential election using twitter data? an integrative modelling approach. *Ann. GIS* **27**, 43–56 (2021).

Acknowledgements

We thank Manlio De Domenico and Michelangelo Puliga for useful discussions. G.C. acknowledges EU project NODES and the partial support by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU.

Author contributions

All the authors conceived the experiment and analysed the results. All authors reviewed the manuscript. The 2022 Twitter data were collected and made available by M.D.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69354-y>.

Correspondence and requests for materials should be addressed to L.F.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access 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/>.

© The Author(s) 2024