Nonlinear Dynamical Analysis of Twitter Time Series

Andrey Dmitriev, Vitaly Silchev, Victor Dmitriev, and Svetlana Maltseva
National Research University Higher School of Economics, Moscow, Russia
a.dmitriev@hse.ru

Abstract. In this paper we present the results of nonlinear dynamical analysis of Twitter time series. According to these results we compare nonlinear dynamical model and nonlinear random dynamical model of Twitter with observed data. From results of nonlinear analysis if observed Twitter time series and evaluation of their probability density functions we conclude, that the most adequate forecasting model of social network is nonlinear random dynamical system. We determine that observed TTS have q-exponential distribution with $1/f^\beta$ noise. Also we consider possible applications of Tsallis entropy and self-organized criticality for analysis of Twitter.

Keywords: Twitter time series-Fractal dimensions-q-exponential distribution-1/f noise-Nonlinear dynamical system-Nonlinear random dynamical system

1 Introduction

Microblogging is one of the most important instruments of business development nowadays. It is actively used for promotion of goods or services, making the positive opinion about the company and allows organizing and supporting customer relationships processes. Corporate microblogging networks and services serve as a platform for business communications between the employees in companies on different scales.

Modeling of processes taking place in microblogging social networks (one of the well-known examples is Twitter) is a complicated, but at the same time theoretically and practically important scientific problem. Results and conclusions that can be made by using such models allow us to identify whether the social network is able to remain stable under the internal and external informational influence, to define different ways of local community formation and to find out the parametric terms of social network management. Such modeling may have a large variety of practical applications. Thus, it can be useful for decision-making processes during the development of short-term and long-term marketing strategies, development of recommender systems, demand forecasting, as well as tasks related to the national security.

There are a number of works in the field of physical modeling of social networks. The main physical models of the social networks are following: Ising model [1-3], Bose-Einstein condensate model [4, 5], quantum walk model [6], ground state and community detection[7], etc. The other relevant works in this area are those of refs. [8-12].
The weak point of the observed papers is that they do not cover nonlinear dynamical analysis of aggregated twitter time series\(^1\) (TTS). Results of such analysis can provide a possibility to select the most appropriate prediction methods for TTS and give a general idea about adequate models of social networks generating these signals.

Recently, more attention has been paid to the study of time series from the point of view of chaos theory. Research in this direction will reveal the nature and interconnections between the hidden processes occurring in microblogging social networks, which will enable the construction of more adequate forecasting models for TTS and a deeper understanding of social networks functioning.

Analysis of chaotic phenomena requires methods and techniques for identifying of time series that is chaotic or having a chaotic component, as well as for quantitative evaluation of chaotic characteristics and comparison of theoretical and experimental time series. Having these methods and techniques allows one to answer the following problems: 1) the number of variables essential for modeling of system dynamics; 2) relation between changes in characteristics and changes in dynamical behavior of the system.

These methods and techniques are grouped into two different, but connected approaches. The first approach focuses on dynamical characteristics of chaos: the Lyapunov exponents and entropy measures, power spectral density and autocorrelation function. The second approach represents the geometric nature of trajectories in the state space considering fractal and correlation dimensions.

These two approaches complete each other. It is intuitively expected that they are closely interconnected. However, theoretical proof of such connection has not been developed yet. That is why we used several criteria of chaotic nature of time series.

This paper is organized as follows. In section 2 we present the results of fractal analysis for empirical TTS with their interpretation. In section 3 we present the results of fractal analysis and probability density function (PDF) for a sample of 3-dimensional nonlinear dynamical model of Twitter network as an open nonequilibrium system\([13]\), as well as comparison with empirical results. In section 4 we provide the results of fractal analysis and PDF for the model of Twitter network as nonlinear random dynamical system comparing them with empirical results and describe the possibilities of applying the Tsallis entropy and self-organized criticality for analysis of TTS. Section 5 contains the conclusions of this paper.

2 Analysis of an Empirical Twitter Time Series

For analysis of empirical TTS we choose the following time series obtained from the resource Mozdeh "BigDataTextAnalysis" (http://mozdeh.wlv.ac.uk):

- bbc_breaking, from 16/05/29 to 17/05/26, step 1 hour;
- cnn_breaking, from 16/07/12 to 17/01/11, step 1 hour;
- nasa, from 16/09/26 to 17/05/26, step 1 hour.

\(^1\) Series of tweet and retweet numbers indexed in time order, \(TR_t\).
Figure 1 shows the corresponding time series.

![Figure 1](image)

**Fig. 1.** Twitter time series: (a) bbc_breaking, (b) cnn-breaking, (c) nasa

It is clear that these time series represent impulse-type signals with integer values. The nonlinear analysis was conducted for all chosen TTS. Such measures as correlation dimension ($D_2$), embedding dimension ($m$), Hurst exponent ($H$) and fractal dimension ($D_F$) were calculated (table 1).

<table>
<thead>
<tr>
<th>Time series</th>
<th>$D_2$</th>
<th>$m$</th>
<th>$H$</th>
<th>$D_F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbc_breaking</td>
<td>3.732</td>
<td>6</td>
<td>0.7648</td>
<td>1.2352</td>
</tr>
<tr>
<td>cnn_breaking</td>
<td>3.984</td>
<td>6</td>
<td>0.8165</td>
<td>1.1835</td>
</tr>
<tr>
<td>nasa</td>
<td>4.202</td>
<td>6</td>
<td>0.7833</td>
<td>1.2167</td>
</tr>
<tr>
<td>Dynamical system</td>
<td>1.896</td>
<td>3</td>
<td>0.5328</td>
<td>1.4272</td>
</tr>
<tr>
<td>Random dynamical system</td>
<td>4.619</td>
<td>5</td>
<td>0.7872</td>
<td>1.2128</td>
</tr>
</tbody>
</table>

The determination of the correlation dimension [14] for a supposed chaotic process directly from experimental time series is often used to gain information about the nature of the underlying dynamics (see, for example, contributions in ref. [15]. In particular, such analysis has been made to support the hypothesis that the time series are generated from the inherently low-dimensional chaotic process [15]. The geome-
try of chaotic attractors can be complex and difficult to describe. It is therefore useful to understand quantitative characterizations of such geometrical objects. One of these characterizations is $D_2$. $D_2$ has several advantages in comparison to the other dimensional measures:

- $D_2$ is easy to compute from the TTS;
- If $D_2$ is finite, then the TTS is a chaotic time series (generated by a dynamical system);
- If $D_2 \rightarrow \infty$, then the TTS is a stochastic time series (generated by a purely random process).

The correlation dimension of the attractor of dynamical system can be estimated using the Grassberger–Procaccia algorithm [14].

For calculation of $D_F$ we used the algorithm, described in a paper [16]. If $D_F > d_T$ ($d_T$ is a topological dimension of the TTS, that equals 1 for all time series), then the TTS is a random fractal. A value of $H = 2 - D_F$ characterizes the following features of the TTS:

- If $H > 0.5$, then the TTS represents a persistent process (a positive increment of a number of tweets and retweets in the past on the average means that there is a tendency to further increase in future, and vice versa);
- If $H < 0.5$, then the TTS represents an anti-persistent process (a positive increment in a number of tweets and retweets in the past on the average means that there is a tendency to decrease in future, and vice versa);
- If $H = 0.5$, then the TTS represents an intermediate state between the persistent and anti-persistent processes (the TTS is a stochastic time series).

In addition, the value of $H$ allows to give a noise classification ($1/f$-classification, where $f$ is a signal frequency) of the TTS [17]:

- If $0 < H \leq 0.5$, then the TTS represents a process with the negative memory, $1/f$ noise or a pink noise (if there has been the positive increment in a number of tweets or retweets, then there is a high probability of appearance of the negative increment in future, and vice versa);
- If $0.5 < H \leq 1$, then the TTS represents a process with a positive memory, $1/f^\beta$ ($\beta > 2$) noise or a brown noise (if there has been the positive increment in a number of tweets or retweets, then there is a high probability of appearance of the positive increment in future, and vice versa);
- If $H = 0.5$, then TTS represents a process with the absence of memory, $1/f^2$ noise or brown noise (the next increment in the number of tweets and retweets doesn’t depend on the previous increments).

Thus, according to the point values of measures, shown in a table 1, the following conclusions can be made:

- TTS is a chaotic time series, i.e. it is generated by dynamical systems in a phase space dimension that equals 6;
• TTS has a fractal structure;
• TTS represents processes with the positive memory;
• TTS represents the persistent process;
• TTS is a signal with the $1/f^\beta$ noise (in support of that, fig. 2 provides the spectral power density plots in log-log scale for corresponding TTS).

![Fig. 2. Power spectral density for TTS: (a) bbs_breaking, (b) cnn_breaking, (c) nasa](image)

3 Twitter Time Series as a Realization of the Nonlinear Dynamical System

Paper [13] proposes a model of Twitter social network as an open nonequilibrium system. Omitting the detailed construction of dynamical system, the model of Twitter is described by well-known Lorenz–Haken equations:

$$\dot{x}_1 = -\alpha x_1 + \beta x_2, \quad \dot{x}_2 = -\gamma x_2 + cx_2x_3, \quad \dot{x}_3 = \epsilon (I_0 - x_3) + k x_1 x_2$$

In equation (1) $x = TR(t) - TR_{eq}$ represents the scaled deviation of number of tweets and retweets ($TR(t)$) from equilibrium value $TR_{eq}$; $x_2(t) = I(t) - I_{eq}$ is the scaled deviation of aggregated internal amount of information ($I(t)$) from equilibrium value $I_{eq}$; $x_3(t) = N_{[u]}(t) - N_{[l]}(t)$ is instantaneous difference in number of users between state $[u]$ and state $[l]$. According to the model, a particular user, being $[u]$-state, has enough information for sending tweet or retweet. If the user is in $[l]$-state (so, he...
or she does not have enough amount of information), then he or she will not send any
tweets or retweets. Control parameter $I_0$ is the intensity of external information flow.

The most important conclusions from model implementation are: 1) impossibility of
social network being in equilibrium state and occurrence of low-dimensional chaos
[18] in social network for specific values of $I_0$. Fig. 3 as hows integral trajectory of
dynamical system (1) ($x_1(t)$), demonstrating the existence of chaotic dynamics in
case of significant intensity of external information flow.

![Fig. 3. Integral trajectory (a) and its histogram (b)](image)

Except for values of higher Lyapunov exponent [19] as one of the measures of
low-dimensional chaos, paper [13] does not contain calculated fractal dimensions for
observed TTS.

Estimations of measures of the chaos for theoretical TTS (fig. 3a). Table 1 contains
the estimated values for measures of chaos for the theoretical TTS (see dynamical
system). Thus, 3-dimensional dynamical model of Twitter as open nonequilibrium system [13]
explains some properties of social network functioning such as fractality,
chaotic nature, persistency and positive memory of TTS.

The weakness of this model lies in significant discrepancy between empirical (fig.
1) and theoretical (fig. 3a) trajectories of TTS. Moreover, it is impossible to fit theo-
retical trajectories to observed data by varying control parameters (in range of chaotic
state) of dynamical system [13]. As it shown on fig. 3b, this dynamical system has 3
stable equilibrium points (three maxima of the histogram) for any values of control
parameters in range of chaotic state.

There are at least two possible ways to achieve the fitness between empirical and
theoretical TTS: by adding specific noise to dynamical system [13] or by using one-
dimensional nonlinear random dynamical system [20] as a model of Twitter network.
According to table 1 at $n = 6$ the estimated value of correlation dimension reaches its
"saturation point" and stops changing significantly. Because of that, the actual number
of variables for constructing an adequate model is 6, but not 3 as it is for model [13].
We do not rule out, that six-dimensional model of Twitter network could explain
existing experimental characteristics, including empirical PDF of Twitter time series.
4 Twitter Time Series as a Realization of the Nonlinear Random Dynamical System

In such autonomous dynamical systems as $\dot{X} = F(X)$, low-dimensional chaos can appear only at $n \geq 3$ [18]. Therefore, the one of opportunity to build an adequate model of a microblogging network is to consider it as a random dynamical system (RDS). In this case, the observable TTS is one of the realizations of $x(t)$ of a stochastic differential equation of the following kind:

$$dx = f(x, t)dt + g(x, t)dW$$

where $W(t)$ is a standard Wiener process.

One of the ways to solve the equation (2) is to find its solution in a form of a probability density function (PDF) $p(x, t)$. In this case, the equation (2) can be transformed into the Fokker-Planck equation [21], that represents a differential equation in partial derivatives of the following kind:

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x}(f(x)p(x, t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(g^2(x)p(x, t))$$

(3)

In this case it is necessary to define the PDF for the empirical TTS (a stationary solution of (3)). Having found out the explicit kind of PDF, we shall be able to find out the explicit kind of (1), describing the realizations of the empirical TTS.

Figure 4 provides PDFs for empirical TTS, which form point to the fact that it is $q$-exponential distribution [22-24]:

$$p(x) = (2 - q)\lambda \exp_q(-\lambda x)$$

(4)

where $\exp_q(x) = [1 + (1 - q)x]^{1/(1-q)}$. 

![Graph showing PDFs for empirical TTS](image-url)
The distribution (4) is a two-parameter generalization ($q < 2$ is a shape parameter, $\lambda > 0$ is a rate parameter) of a one-parameter exponential distribution. Table 2 contains the estimated values for parameters of PDF (4) obtained by maximum likelihood method [25].

**Table 2.** Point and interval estimations of the PDF (4) parameters

<table>
<thead>
<tr>
<th>User</th>
<th>$q$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbc_breaking</td>
<td>1.202±0.005</td>
<td>1.980±0.074</td>
</tr>
<tr>
<td>cnn_breaking</td>
<td>1.155±0.025</td>
<td>1.482±0.086</td>
</tr>
<tr>
<td>nasa</td>
<td>1.184±0.038</td>
<td>1.362±0.069</td>
</tr>
</tbody>
</table>

From table 2 we conclude that empirical PDF corresponds to $q$-exponential distribution.

Going back to the equation (2): a stationary probability density function of the TTS looks as (4) with the numerical parameter values shown in a table 2 and is a stationary solution of the equation (3). Therefore, the equation (3) should be of such kind, that gives the distribution (4) for all realizations of the random dynamical system.

A group of researchers [26-28] has suggested the RDS in a view of a nonlinear stochastic differential equation:

$$dx = \sigma^2 \left( \eta - \frac{1}{2} \lambda \right) (x + x_0)^{2\eta-1} dt + \sigma (x + x_0)^\eta dW$$

(5)

where $x(t) \geq 0$ is a signal, $\eta \neq 1$ is a power-law exponent of the multiplicative noise, $\lambda > 0$ is a parameter, defining the behavior of stationary probability distribution, $W$ is a standard Wiener process, $\sigma$ is a parameter of the multiplicative noise. Parameter $x_0$ limits the divergence of the power-series distribution $x(t)$ by $x(t) \to 0$. If $x \ll x_0$, then (5) generates a linear additive stochastic process (Brownian movement with the stable drift); if $x \gg x_0$, then (5) generates a multiplicative process [27].

If $x_0 = 1$, then the stationary solution of the equation (3) takes the form of an $q$-exponential distribution (4) by $q = 1 + 1/\lambda$. Besides, some of realizations of the process (5) give a power spectral density in a form of $1/f^\beta$.
We have calculated estimations of the measures of chaos for some realizations of RDS (5). Table 1 contains the estimated values for measures of chaos for the theoretical TTS (see random dynamical system).

Thus, the realizations of the RDS (5) have not only close measures to the observable fractal measures of the TTS (table 1) in comparison to the realizations of the dynamical system [13], but they also have an observable (table 2) $q$-exponential distribution. Therefore, the RDS (5) is more adequate model in comparison to the model in a form of the dynamical system [13].

$q$-exponential distribution takes place by the maximization of the Tsallis entropy [29] considering definite limitations. Tsallis entropy as a non-additive generalization of the Boltzmann-Gibbs entropy has the following form:

$$T_q = \frac{1}{q-1} \left(1 - \sum_{i=1}^{N} p_i^q \right) \quad (6)$$

The probability $p_i = N_i / N(\varepsilon)$ can be estimated in much the same way as that one used in the Renyi entropy; $N_i$ is a number of system elements for the $i$-element of the $\varepsilon$-partition; $N(\varepsilon)$ – is a full number of elements of the given $\varepsilon$-cover. If $q \to 1$, then the entropy (6) transforms into the well-known Shannon entropy.

In contrast to all entropy types, the Tsallis entropy is nonadditive. Being applied to the microblogging network (such as, for example, Twitter) it gives a possibility to correctly describe a social network, where any user interacts not only with the nearest user or several nearest users, but also with the whole network or some of its parts. Besides, from (5) it follows that $T_q$ is concave by $q > 0$ and convex by $q < 0$.

Thus, entropy description of Twitter based on Tsallis statistics is appropriate for studying of evolution of social network that contains large amount of users who interact with each other in a particular way and, specifically, every user can interact not only with his or her nearest neighbors but also with remote users.

There are a lot of practical application of Tsallis theory. Among them there are studies on the anomalous diffusion [30, 31], uniqueness theorem [32], sensitivity to initial conditions and entropy production at the edge of chaos [33] and many others (see ref. [34]).

The fact, that the RDS (5) generates a signal with the power-series distribution (4) and with the occurrence of the $1/f^\beta$ noise [35], is the important feature of the RDS (5). It is determined by the existence of the degree $2\eta - 1$ in the drift term and degree $\eta$ in the noise term. The same fact is observable for the empirical TTS as well.

The existence of the power laws of signal distribution with the presence of the $1/f^\beta$ noises (see fig. 2) is a necessary condition of system complexity, its nontrivial behavior or presence of the catastrophic events (unexpected and/or extraordinary). There is a relatively new field in non-linear dynamics – a theory of the self-organized criticality [36]. It was created to explain similar phenomena in systems with the power-series distributions and $1/f^\beta$ noises.

The existence of the $1/f^\beta$noise in a system means the internal tendency to the catastrophic cases in a system. The theory of the self-organized criticality studies the dynamical dissipative systems with the high range of discretion, which operate in the neighborhood of the critical point without the smallest external influence. If the sys-
tem is in a critical configuration, than small fluctuations can lead to a random event of any “size” with the power-series distribution similar to (4):

\[ p(s) \sim s^{-\tau} \]  

(7)

Twitter as a self-organizing system generates signals with \(1/f\) noise, since the lifetime of events is related to its scale according to [36]:

\[ t^{1+\gamma} \approx s \]  

(8)

where \(\gamma\) is the speed of event distribution in the system.

5 Conclusion

The main contributions of the present paper look as follows:

- The three-dimensional model of the microblogging network [13] (such as, for example, Twitter) as an open non-equilibrium system explains some features of social networks functionality, such as the fractality, chaotic state, persistence, as well as the positive memory of the TTS. But, at the same time, the dimension test of such dynamical system gives the negative result: empirical embedding dimension of all TTS equals to 6 (by \(n = 6\) the correlation dimension reaches the saturation and stops changing). This fact leads to the necessity of building a new model of a microblogging network in a form of nonlinear RDS.

- We have conducted a research into the empirical PDF of some TTS to build a model of the microblogging network in a form of one-dimensional non-linear RDS. As a result it has been recognized that at the significance level equal to 0.05 the observable PDF has a \(q\)-exponential distribution. For such distribution, the one-dimensional nonlinear RDS has been suggested. The fractal measures of its realizations are equivalent to the measures of the observable TTS.

- It has been shown, that in contrast to all entropy types, the Tsallis entropy gives a possibility to correctly describe a network, where any user interacts not only with the nearest user or several nearest users, but also with the whole network or some of its parts. Use of the Tsallis entropy also allows to describe the macroscopic stability of a microblogging network.

- It has also been mentioned, that because of the existence of the \(1/f^\beta\) noise and power series distribution, a social network may have a tendency to catastrophic events. If a social network keeps staying in a critical configuration, then small fluctuations may lead to the random event of any scale.

Despite the fact, that the results of the present study can be useful for the research into the fundamentals of the network functionality, we haven’t yet defined the physical meaning of parameters of the one-dimensional nonlinear RDS. That is the question of our further research.
References


