Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 122 (2017) 812–819

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2017)

Modeling of Microblogging Social Networks: Dynamical System vs. Random Dynamical System

Andrey Dmitriev*, Olga Tsukanova, Svetlana Maltseva

National Research University Higher School of Economics, 33 Kirpichnaya St., 105187 Moscow, Russia

Abstract

Building of adequate dynamical models of microblogging social networks is a topical task that is of interest from both theoretical and practical aspects. Experimental and theoretical results of studies related to choice of the adequate model are presented. The choice was made between two models: a nonlinear dynamical system and a nonlinear random dynamical system. By results of the fractal analysis of observable network time series and defining their probability density function it was established that the nonlinear random dynamical system was more adequate than the nonlinear dynamical system. The character of the observable time series was also explored. The possibility that microblogging social networks can be analyzed by means of Tsallis entropy and self-organized criticality is examined.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Keywords: social network, dynamical system, random dynamical system, time series.

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 various 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 means of using such models allow to identify if the social network is able to stay 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

* Corresponding author. Tel.: +7-495-771-3238; fax: +7-495-771-3238.
E-mail address: a.dmitriev@hse.ru.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.
10.1016/j.procs.2017.11.441

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 is a number of scientific papers in the field of physical modeling of social networks. The main physical models of the social networks are following: Ising model [5,7,12], Bose-Einstein condensate model [2,6], quantum walk model [11], ground state and community detection models [23], etc. The other relevant works in this area are those of refs. [10,21,27-29].

The present paper is the further development of the earlier results. As a matter of fact earlier it was shown that a microblogging network has a fractal structure. Theoretical results of that paper included the proof of existence of the control parameter (the rate of the external data flow); inability of a social network to keep staying in an equilibrium state; and appearance of the low-dimensional chaos [16] in the network at the definite value of the control parameter. With the exception of the values of largest Lyapunov exponent as a measure of existence of low-dimensional chaos, such measures as fractal dimensions for the observed time series in the microblogging networks (values in the time series are the number of tweets and retweets) haven't been calculated.

Thus, the statement that the microblogging networks have the fractal structure requires experimental demonstration by means of testing the values of various fractal dimensions of the time series in the microblogging networks. Except the fractal structure of the social network, the fine structure of theoretical and observable time series should be equivalent by the use of probability density function for a signal. In case of the negative examination of the fractality or the mismatch of the theoretical and observable fine structures of the time series, a new refined microblogging network model (that is relevant to the experimental data) is required. And that is the aim of the present study.

2. Fractal Analysis of Time Series in Microblogging Networks

The present research is based on the analysis of time series in Twitter (one of the most famous microblogging networks). For the fractal analysis of empirical Twitter time series (TTS) some of the time series reflecting the enterprise behavior were chosen (a source: Mozdeh Big Data Text Analysis (http://mozdeh.wlv.ac.uk/). Let us consider them as A, B, C, D:

- A: from 16/10/24 to 17/03/07, step: 1 hour;
- B: from 16/10/25 to 17/03/07, step: 1 hour;
- C: from 16/10/24 to 17/03/07, step: 1 hour;
- D: from 16/07/12 to 17/01/11, step: 1 hour.

The TTS for D is shown at the Fig. 1. The time step is shown at the horizontal axis, the sum of tweets and retweets is shown at the vertical axis.

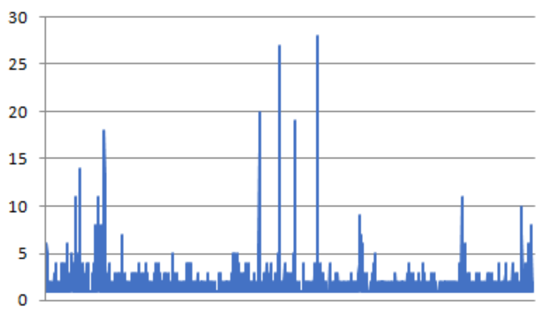

Fig. 1. TTS (D)

The fractal analysis was conducted for all chosen TTS. Such measures as correlation dimension (D_2), correlation entropy (K_2), Hurst exponent (H) and fractal dimension (D_F) were calculated (table 1).

The determination of the correlation dimension [13-14] 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. [8]. In particular, such analysis has been made to support the hypothesis that the time series are generated from the inherently low-dimensional chaotic process [8]. The geometry 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 $D_2 \to \infty$, 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 [13-14]. This algorithm allows to estimate a phase space dimension $n \leq 2D_2 + 1$, considering that the TTS is generated by the dynamical system.

For calculation of D_F we used the algorithm, described in a paper [9]. 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);
- If $H = 1$, then the TTS is a chaotic 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 [20].

- 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).

Correlation entropy K_2 was calculated using the algorithm, which had been presented in a paper [13]. The value of K_2 characterizes the following features of TTS:

- If $K_2 \neq 0$, then the TTS is a chaotic time series;
- If $K_2 \to \infty$, then the TTS is a stochastic time series.

Table 1. Point and interval estimates of the fractal measures for TTS

	D_2	n	H	D_F	K_2
A	0.143	1	0.5548±0.2353	1.4452±0.2353	0.339
B	0.000	1	0.7865±0.1709	1.2135±0.1709	0.098
C	0.000	1	0.4402±0.1588	1.5598±0.1588	0.080
D	0.000	1	0.5733±0.2076	1.4267±0.2076	0.037

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 1;
- TTS has a fractal structure;
- TTS represents processes with the negative or positive memory;
- TTS is a signal with $1/f^{\beta}$ noise.

The numerical value of K_2 allows getting the estimation of the time of predictability for the TTS: $t_F \sim 1/K_2$. Thus, users A, B, C and D can get a forecast from 2 and up to 27 hours. In some situations this time will be enough to make a decision related to the further behavior (such as, for example, change of the current trend) based on the forecasting of the TTS. Such forecast can be improved in case of the sentiment content.

It is important that these conclusions are based on the point estimates of the fractal measures of the TTS, i.e. in average.

Now we are returning to the dynamical model of a microblogging system. The microblogging system is being modelled by the Lorenz-Haken non-linear dynamical system of the third order [15]. Point and interval estimates of the fractal measures for a theoretical TTS (one of the trajectories of the dynamical system) have the following values: $D_2 = 1.896$, $n = 3$, $H = 0.5328 \pm 0.1496$, $D_F = 1.4272 \pm 0.1496$ and $K_2 = 0.549$ [19]. Therefore, the three-dimensional model of the microblogging network as an open non-equilibrium system explains some features of social networks functionality, such as its fractality, randomness (chaotic character), persistence, as well as the positive memory of the TTS. But, in our opinion, the three-dimensionality can be a disadvantage for this model. Actually, according to the table 1 ($n = 1$ for all TTS), i.e. at $n = 1$ the correlation dimension reaches the saturation point and stops changing. Thus, a number of independent dynamic variables of the adequate model equal 1, and not 3, as according to the Lorenz-Haken equation. It is not improbable, that the one-dimensional model of the microblogging network will allow to describe several other characterizations besides the present experimental data (eg., an empirical probability density function of the TTS).

3. Microblogging Network as a Random Dynamical System

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

$$dx = f(x)dt + g(x)dW, \tag{1}$$

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

One of the ways to solve the equation (1) is to find its solution in a form of a probability density function (PDF) $p(x,t)$. In this case, the equation (1) can be transformed into the Fokker-Planck equation, that represents a differential equation in partial derivatives of the following kind:

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

In this case it is necessary to define the PDF for the empirical TTS (a stationary solution of (2)). Having found out the explicit kind of the PDF, we shall be able to find out the explicit kind of (1), describing the realizations of the empirical TTS.

The fig. 2. shows an empirical PDF in a form of a bar chart for the TTS of the user D. A number of tweets and retweets, falling into the intervals, determined by the Sturges rule, are shown at the vertical axis. The empirical PDF for the other users have similar visualizations.

A kind and a character of the empirical TTS and PDF point to the fact, that the PDF has a q-exponential distribution [31, 33]:

$$p(x) = (2-q)\lambda \exp_q(-\lambda x)\,, \quad (3)$$

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

The distribution (3) is a two-parameter generalization ($q < 2$ is a shape parameter, $\lambda > 0$ is a rate parameter) of a one-parameter exponential distribution.

The numerical values of parameters of the distribution (3), derived by the least-squares method, as well as the results of check of a static hypothesis about the equivalence of the empirical PDF and the theoretical density of distribution (a Pirson's χ^2 test has been used in this case as the most frequently used criteria for the check of hypothesis about the belonging of an observable sample to some theoretical distribution law) are shown in a table 2. The following hypothesis H_0 has been checked: $PDF = p(x, q, \lambda)$.

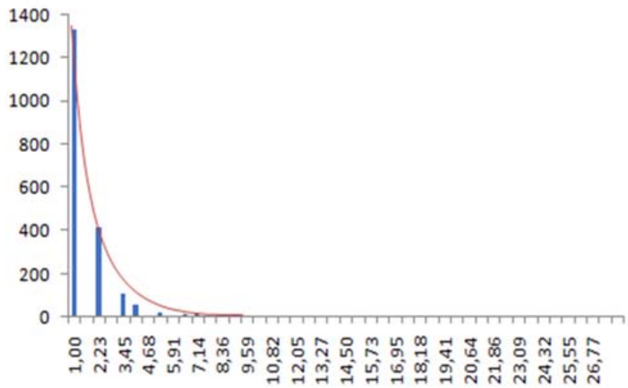

Fig. 2. A PDF for the TTS (D). The red line is an empirical PDF with the parameters $q = 0.662$ and $\lambda = 0.507$

Table 2. Point estimations of parameters of distribution (3), observable (empirical) χ^2 and theoretical $\chi^2_{0.05,m}$ criteria values

	q	λ	χ^2	$\chi^2_{0.05,m}$
A	0.834	0.887	28.532	45.558
B	0.935	0.834	32.431	44.181
C	0.902	0.654	36.834	45.558
D	0.662	0.507	34.536	42.795

According to the table 2, in any case, the hypothesis about the belonging of the observable sample of the PDF to the theoretical distribution law (3).

Going back to the equation (1): a stationary probability density function of the TTS looks as (3) with the numerical parameter values shown in a table 2 and is a stationary solution of the equation (2). Therefore, the equation (1) should be of such kind, that gives the distribution (3) for all realizations of the random dynamical system.

A group of researchers [18, 25-26] has suggested a random dynamical system (RDS) in a view of a nonlinear stochastic differential equation:

$$dx = \sigma^2 \left(\eta - \frac{1}{2}\lambda\right)(x + x_0)^{2\eta-1}dt + \sigma(x + x_0)^{\eta} dW \quad (4)$$

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, σ 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 (4) generates a linear additive stochastic process (Brownian movement with the stable drift); if $x \gg x_0$, then (4) generates a multiplicative process [25].

If $x_0 = 1$, then the stationary solution of the equation (2) takes the form of the q-exponential distribution (3) by $q = 1 + 1/\lambda$. Besides, some of realizations of the process (4) give a power spectral density in a form of $1/f^{\beta}$.

We have calculated point and interval estimations of the fractal measures for some realizations of RDS (4): $D_2 = 0.234$, $n = 1$, $H = 0.6844 \pm 0.1206$, $D_F = 1.3156 \pm 0.1206$ and $K_2 = 0.228$.

Thus, the realizations of the RDS (4) have not only close measures to the observable fractal measures of the TTS (table 1) in comparison to the realizations of the Lorenz-Haken dynamical system, but they also have the observable (table 2) q -exponential distribution. Therefore, the RDS (4) is more adequate model in comparison to the model in a form of the Lorenz-Haken dynamical system.

q-exponential distribution takes place by the maximization of the Tsallis entropy [30] considering definite limitations. The Tsallis entropy as a non-additive generalization of the Boltzmann-Gibbs entropy has the following form:

$$T_q = \frac{1}{q-1}\left(1 - \sum_{i=1}^{N} p_i^q\right);\ \sum_{i=1}^{N} p_i = 1, q \in \mathbb{R} \quad (5)$$

The probability $p_i = N_i/N(\varepsilon)$ can be estimated in much the same way as that one used in the Renyi entropy [24]: N_i is a number of system elements for the i- element of the ε-partition; $N(\varepsilon)$ – is a full number of elements of the given ε-cover. If $q \to 1$, then the entropy (5) 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 $q < 0$.

There are a lot of practical application of Tsallis theory. Among them there are studies on the anomalous diffusion [22,34], uniqueness theorem [1], sensitivity to initial conditions and entropy production at the edge of chaos [4] and many others (see ref. [32]).

The fact, that the RDS (4) generates a signal with the power-series distribution (3) and with the occurrence of $1/f^{\beta}$ noise [17] is the important feature of the RDS (4). It is determined by the existence of the degree $2\eta - 1$ in the drift term and degree η 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 $1/f^{\beta}$ noises 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 self-organized criticality [3]. It was created to explain similar phenomena in systems with power-series distributions and $1/f^{\beta}$ noises.

The existence of $1/f^{\beta}$ noise in a system means the internal tendency to the catastrophic cases in the 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 system is in a critical configuration, than small fluctuations can lead to a random event of any "size" with the power-series distribution similar to (3).

4. Conclusion

The main contributions of the present paper look as follows:

- The three-dimensional model of the microblogging network (such as, for example, Twitter) as an open non-equilibrium system explains some features of social networks functionality, such as fractality, randomness (chaotic character), 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 dimension of all TTS equals to 1 (by $n = 1$ 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 one-dimensional 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 the 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 the Tsallis entropy, in contrast to all entropy types, 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 $1/f^{\beta}$ noise and a 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 further research.

References

[1] Abe S. Axioms and uniqueness theorem for Tsallis entropy. Physics Letters A (271) ; 2000.
[2] Albert R., Barabási A.-L. Statistical mechanics of complex networks. Rev. Mod. Phys. (74); 2002.
[3] Bak P., Tang C., Wiesenfeld K. Self-organized criticality: an explanation of 1/f noise. Physical Review Letters (59); 1987.
[4] Baldovin F., Robledo A. Nonextensive Pesin identity: Exact renormalization group analytical results for the dynamics at the edge of chaos of the logistic map. Physical Review E (69); 2004.
[5] Bianconi G. Mean field solution of the Ising model on a Barabasi-Albert network. Phys. Lett. A (303); 2002.
[6] Bianconi G., Barabási A.-L. Bose-Einstein Condensation in Complex Networks. Phys. Rev. Lett. (86); 2001.
[7] Dasgupta S., Pan R. K., Sinha S. Phase of Ising spins on modular networks analogous to social polarization. Physical Review E (80); 2009.

[8] Ding M., Grebogi C., Ott E., Sauer T., Yorke J. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? Physica D (69); 1993.
[9] Dubovikov M. M., Starchenko N. S., Dubovikov M. S. Dimension of the minimal cover and fractal analysis of time series. Physica A (339); 2004.
[10] Ebel H., Davidsen J., Bornholdt S. Dynamics of Social Networks. Complexity (8); 2003.
[11] Faccin M., Johnson T., Biamonte J., Kais S. Degree Distribution in Quantum Walks on Complex Networks. Phys. Rev. (3); 2013.
[12] Grabowski A., Kosinski, R. A. Ising-based model of opinion formation in a complex network of interpersonal interactions. Physica A (361); 2006.
[13] Grassberger P., Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal. Phys. Rev. A (28); 1983.
[14] Grassberger P., Procaccia I. Measuring the trangeness of strange attractors," Physica D: Nonlinear Phenomena (9); 1983.
[15] Haken H. Analogy between Higher Instabilities in Fluids and Lasers. Phys. Lett. A (53); 1975.
[16] Hilborn R.C. Chaos and nonlinear dynamics: an introduction for scientists and engineers, United Kingdom: Oxford University Press; 2000.
[17] Kaulakys B., Alaburda M. Modeling scaled processes and $1/f^{\beta}$ noise using nonlinear stochastic differential equations. J. Stat. Mech. 2009(b).
[18] Kaulakys B., Alaburda M., Gontis V., Ruseckas J. Modeling long-memory processes by stochastic difference equations and superstatistical approach. Brazilian Journal of Physics (39); 2009(a).
[19] Lichtenberg A., Lieberman M. Regular and Stochastic Motion, New York: Springer-Verlag; 1983.
[20] Mandelbrot B. B. , Ness V. Fractional Brownian motions, fractional noises and aplications. SIAM Rev. (10); 1968.
[21] Mendes V. Tools for network dynamics. J. Bifurcation Chaos (15); 2005.
[22] Plastino A. R., Plastino A. Non-extensive statistical mechanics and generalized Fokker-Planck equation. Physica A (222), 1995.
[23] Reichardt J., Bornholdt S. Statistical mechanics of community detection. Phys. Rev. E (74); 2006.
[24] Renyi A. On measures of information and entropy. Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley, Los Angeles; 1961.
[25] Ruseckas J., Kaulakys B. Tsallis distributions and 1/f noise from nonlinear stochastic differential equations. Phys. Rev. E (84); 2011.
[26] Ruseckas J., Gontis V., Kaulakys B. Nonextensive statistical mechanics distributions and dynamics of financial observables from the nonlinear stochastic differential equations. Advances in Complex Systems (15); 2012.
[27] Skaza J., Blais B. Modeling the infectiousness of Twitter hashtags. Physica A (465); 2017.
[28] Toivonen R., Onnela J.-P., Saramaki J., Hyvonen J., Kaski K. A model for social networks. Physica A (371); 2006.
[29] Toivonen R., Onnela J.-P., Saramaki J., Hyvonen J., Kaski K. A comparative study of social network models: Network evolution models and nodal attribute models. Social Networks (31), ; 2009.
[30] Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics (52); 1988.
[31] Tsallis C. What are the numbers that experiments provide? Quimica Nova. (17), 1994.
[32] Tsallis C. Nonextensive Statistics: Theoretical, Experimental and Computational Evidences and Connections. Brazilian Journal of Physics (29); 1999.
[33] Tsallis C. Nonadditive entropy and nonextensive statistical mechanics-an overview after 20 years. Braz. J. Phys. (39); 2009.
[34] Tsallis C., Bukman D. Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis. Physical Review E (54); 1996.