

## Research Article

# Self-Organized Criticality on Twitter: Phenomenological Theory and Empirical Investigation Based on Data Analysis Results

Andrey Dmitriev <sup>1</sup>, Victor Dmitriev <sup>1</sup>, and Stepan Balybin <sup>2</sup>

<sup>1</sup>School of Business Informatics, National Research University Higher School of Economics, Moscow, Russia

<sup>2</sup>Department of Physics, Lomonosov Moscow State University, Moscow, Russia

Correspondence should be addressed to Andrey Dmitriev; [a.dmitriev@hse.ru](mailto:a.dmitriev@hse.ru)

Received 16 August 2019; Revised 6 November 2019; Accepted 11 December 2019; Published 27 December 2019

Guest Editor: Raúl Baños

Copyright © 2019 Andrey Dmitriev et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, there has been an increasing number of empirical evidence supporting the hypothesis that spread of avalanches of microposts on social networks, such as Twitter, is associated with some sociopolitical events. Typical examples of such events are political elections and protest movements. Inspired by this phenomenon, we built a phenomenological model that describes Twitter's self-organization in a critical state. An external manifestation of this condition is the spread of avalanches of microposts on the network. The model is based on a fractional three-parameter self-organization scheme with stochastic sources. It is shown that the adiabatic mode of self-organization in a critical state is determined by the intensive coordinated action of a relatively small number of network users. To identify the critical states of the network and to verify the model, we have proposed a spectrum of three scaling indicators of the observed time series of microposts.

## 1. Introduction

The general science development trend in the 20th century, which is also passed in the new century, is the gradual penetration of ideas and methods of physics in natural as well as traditional humanities. Since the 1970s, the methods of mathematical and then physical modeling have been increasingly used in such sciences as demography, sociology, economics, history, and political science. In all these sciences, the desire for an objective and, preferably, a quantitative description of various social and economic phenomena is increasing.

The development of quantitative models in sociology, political science, theory of transport flows, and other areas of society investigations is gradually moving relevant tasks from the humanities and engineering sciences to interdisciplinary applications of mathematics and physics. In the literature of recent years, the term sociophysics [1, 2] is assigned to all such areas. The main task of this new field of natural science is to search for objectively measurable and formalizable patterns that determine various social

processes. Sociophysics analyzes the structure and dynamics of all existing varieties of social systems, using ideas and methods borrowed from theoretical and experimental physics.

Some of the objects and phenomena studied by sociophysics are social networks (e.g., see the review [3] and references therein) and critical phenomena, such as phase transitions, observed in them (e.g., see the reviews [4, 5] and references therein). Dorogovtsev and co-authors state in their paper [4] that “Critical phenomena in networks include a wide range of issues: structural changes in networks, the emergence of critical—scale-free—network architectures, various percolation phenomena, epidemic thresholds, phase transitions in cooperative models defined on networks, critical points of diverse optimization problems, transitions in co-evolving couples—a cooperative model and its network substrate, transitions between different regimes in processes taking place on networks, and many others.” In the thermodynamics theory of irreversible processes, it is stated that significant structure reconstructions occur when the external parameter reaches a certain critical value (the

bifurcation point) and has the character of a kinetic phase transition [6]. The critical point is reached as a result of fine-tuning of the system external parameters. In a certain sense, such critical phenomena are not robust. The following types of self-organization of nonlinear systems can be identified that lead to nonrobust critical phenomena: self-organization during phase transitions, which are characterized by spatial-temporal scale invariance with a transition to the critical point, when the external parameter reaches its critical value; self-organization during geometric phase transitions, when the critical value of the cell filling probability is reached (for example, the percolation threshold); and self-organization of dissipative structures at the bifurcation point, in case when some external parameter (for example, the temperature gradient in the classical Benard problem) reaches its critical value.

At the end of the 1980s, Bak et al. [7, 8] found that there are complex systems with a large number of degrees of freedom that go into a critical mode as a result of the internal evolutionary trends of these systems. A critical state of such systems does not require fine-tuning of external control parameters and may occur spontaneously. Thus, the theory of self-organized criticality (SOC) was proposed. This is the theory claiming to be the universal theory, explaining the spontaneous occurrence of critical states in open nonequilibrium systems. A characteristic feature that qualitatively distinguishes SOC from other phenomena with a similar nature is the realization of a self-organized critical state in a wide range of external control parameters. For SOC, however, no special “parameter tuning” is required, and, in this sense, such critical phenomena are robust.

From the moment of the SOC model emergence, this model started to be applied to describe critical phenomena in systems regardless of their nature (e.g., see the review [9] and references therein). Not an exception is the application of the theory to the description of critical phenomena in social networks (e.g., see the works [10–13]).

The motivation of our investigation is the following. There is a number of studies (e.g., see the works [11, 13–20]), in which it is established that the observed flows of microposts generated by microblogging social networks (e.g., Twitter) are characterized by avalanche-like behavior. Time series of microposts depicting such streams are the time series with a power-law distribution of probabilities, with  $1/f$  noise and long memory. As it will be shown in Section 2.2, these characteristics of the time series are key features of the social network in the SOC state. Despite this, there are no studies on the construction, analysis, and verification of physical models that explain the phenomenon of the emergence and spread of avalanche of microposts on Twitter. The construction, analysis, and verification of such a phenomenological model are the purpose of our research.

The presented work is structured as follows. Section 2.1 is devoted to the description of one of the scenarios of Twitter self-organizing transition in a critical state, determined by the specifics of the network functioning. Section 2.2 introduces the notion of a spectrum of indicators of self-organized network criticality, which is an identifier that allows

the SOC state of the network to be distinguished from a noncritical state, determined from the results of the analysis of the time series of microposts. Methods of mining and data analysis for identifying the state of the network, as well as the results of identifying network states from the observed data, are presented in Section 3. These data are used to verify the model presented in Section 4, which describes the conditions for Twitter self-organization transition to the critical state from its subcritical state. Section 5 presents the general conclusions from the study, as well as their discussion. Finally, Section 6 is devoted to a discussion of tasks that cannot be solved within the proposed model, and a brief description of the approaches to their solution.

## 2. Self-Organized Criticality on Twitter

This section presents a qualitative nonformalized description of the emergence mechanisms of a self-organized critical state on Twitter as a result of the coordinated action of strategically oriented network users. The range of indicators of self-organized criticality of a social network is defined as the identifier of the network functioning in the subcritical (SubC), the self-organized critical (SOC), and the supercritical (SupC) states.

### 2.1. Mechanisms of Self-Organized Criticality on Twitter.

The well-known physical model of self-organized criticality, Abelian sandpile model [21], provides one of the scenarios for the system to achieve a self-organized critical state in the robust case. A model with a pile of sand is metaphorical, and the real dynamics of such systems can be very diverse. We adapt this scenario to explain the emergence of the SOC state on Twitter, accompanied by the appearance of the avalanche of microposts on the network of various sizes: from small avalanches of the order of 10 microposts per second to large avalanches of about 1000 microposts per second or even more. The corresponding time series shows consistently measured numbers of microposts (avalanche sizes) at some time intervals.

We hypothesize that Twitter self-organization in a critical state results from the consistent behavior of a relatively small number of network users ( $S$ ) when  $S$  reaches a critical state  $S_c$ . Further discussion is devoted to the substantiation of this assumption.

First of all, we introduce the concepts that will be needed for further discussion. Let  $\mathcal{N}$  be the total number of Twitter users who are interested in a particular topic, for example, “2016 United States Presidential Election.” Such users, who are united by an interest in a given topic, form a community on the network and, therefore, only these users can send microposts relevant to this topic. Let that  $S$  ( $S \ll \mathcal{N}$ ) users of this community follow a certain strategy. For example, the goal of these users is to pump the network, i.e., to distribute as many microposts, related to a certain topic on the network, as possible. Call these users as strategically oriented users (SOUs). SOUs have a common goal, to achieve that they act in concert. Consolidation of users into a subcommunity of SOUs does not necessarily occur as a result of

a preliminary conspiracy, but also unconsciously. Examples of concerted action to achieve a common goal can be the promotion of a candidate in the political elections, as well as the coordination of actions of participants in protest movements and/or the involvement of citizens in protest movements. Recent protest movements have suggested that online social networks might play a key role in their organization, as adherents have a fast, many-to-many communication channel to help coordinate their mobilization [22]. The behavior of network users during natural disasters may be the result of their unconscious collusion. The remaining  $\mathcal{N} - S$  users do not follow a single coherent strategy and, in this sense, are randomly oriented users (ROUs).

The rationale for dividing network users into two classes is the research results presented by Pramanik and co-authors in their paper [23]. They introduce two mention strategies: random mention and smart mention to model the mention preferences of the users. They proposed a model of the cascade formation in Twitter, incorporating both retweet and mention activities. Realizations of the model prove the elegance of smart mention strategy in boosting tweet popularity, especially in the low retweeting environment.

We use, in our opinion, the exhaustive classification of Twitter users presented in [24] to classify network users to one of the two classes. The classification consists of real users, which include personal users, professional users, and business users, and also digital actors, which include spam users, feed/news, and viral/marketing services. Of course, the assignment of one of these classes to SOUs or ROUs is conditional and is determined by the specifics of the topic, discussed by network users. However, in most cases, all digital actors who are using bots, professional users, and business users can be considered as SOUs. For example, the main goal of professional users and business users on the network is to involve as many users as possible, for example, personal users, in a discussion of a certain topic. Finally, personal users can be considered as ROUs. Indeed, such users create their Twitter profiles for entertainment, training, or to receive news, etc. This is the most numerous class of Twitter users.

Consider the features of the network users' interactions, which lead to the emergence of the SOC state on Twitter. To explain the mechanisms of the emergence of such state, it is appropriate to distinguish three consecutive network states: the SubC state, the SOC state, and the SupC state.

The SubC state is the chaotic network state, which is observed in a certain time interval or in subcritical time  $\Delta t_{\text{SubC}}$  (from 0 to  $t_c$ ). A demonstration of the chaotic nature of this state is the chaotic (disordered) distribution of microposts avalanches on the network. The most common scenario for observation of this state corresponds to the presence of only ROUs ( $S = 0$ ) on the network, who generates avalanches of microposts in terms of a certain topic. Such avalanches are not interconnected, and they are small in size and quickly fade out in time and space. This is due to the fact that ROUs do not behave in a consistent manner, do not pursue one goal, and do not pump the network with certain information, and, accordingly, it does

not lead to the formation of avalanches of microposts of all sizes. In such a network, self-organization in an ordered state, which is characterized by the existence of avalanches of microposts of all sizes, is impossible. ROUs are not characterized by cooperative (synchronous) behavior and, therefore, a spontaneous transition of the network from a chaotic to an ordered state, in which avalanches of microposts of all sizes distribute on the network, is not possible. We do not exclude that in the community of ROUs, connected exclusively by discussing a certain topic, there are local structures with a small number of hierarchical levels (user, his subscribers, subscribers of their subscribers, etc.). The avalanches follow a single tweet that is retweeted or similar tweets as they move across the network. But the avalanche of microposts distributed in such structure has relatively small size. Even the totality of such hierarchical structures will allow to generate only many avalanches of small sizes that are not interconnected.

Suppose that at each moment of time, one SOU ( $S \neq 0$ ) goes on Twitter, wherein  $S < S_c$ . These users act in concert, trying to form the avalanches of microposts of all sizes relevant to a certain topic. Gradually, SOUs form hierarchical structures on the network. The cooperative behavior of these users gives them the opportunity to build hierarchical structures that are quite effective to generate avalanches of microposts of larger sizes. If SOUs are real users and most of them are influential persons of Twitter, then it is possible to form a hierarchical structure with a large number of levels (influential person #1, his subscribers, subscribers of their subscribers, etc.), which can generate avalanches of microposts of greater sizes. Even larger avalanches can be generated, if both ROUs, which are subscribers of subscribers on certain levels, and other influential persons (influential person #2, influential person #3, etc.) with their subscribers, including SOUs and ROUs, will be integrated in this structure. In some sense, ROUs are an active environment for increasing the size of the avalanches originally generated by SOUs.

It should be noted that the considered hierarchical structure is not the only structure through which it is possible to generate avalanches of microposts of larger sizes. Other possible mechanisms for generating criticality will be described in Conclusion. Nevertheless, the above-mentioned mechanism of the spread of microposts avalanches of all sizes, in our opinion, is the most justified. This is determined by the basic specifics of users' organization on Twitter: user (hierarchical level #1), his subscriber (hierarchical level #2), subscriber of his subscriber (hierarchical level #3), etc. Moreover, in [25], it was shown that the evolution of hierarchically subordinate complex networks reduces to anomalous diffusion in the ultrametric space of the hierarchical system. The stationary distribution over the levels of such a system is determined by a power law. Besides, Bakshy and co-authors state in their paper [26] "Unsurprisingly, we find that the largest cascades tend to be generated by users who have been influential in the past and who have a large number of followers." This study also provides a rationale for the existence of a hierarchical structure (influential person #1, his subscribers, subscribers

of their subscribers, etc.), as the most effective structure for the distribution of large avalanches of microposts. In the paper [27], a conceptual and practical model is proposed for the classification of topical networks on Twitter, based on their network-level structures. The existence of connection between hierarchical sequences of tweet-retweet-follow and cascades of retweets is discussed in [28].

If  $S$  does not reach its critical value  $S_c$ , then unrelated avalanches of microposts, although of larger sizes, are still forming on Twitter. The formed hierarchical system of SOUs and ROUs on the social network is still not able to form an avalanche of microposts of all sizes. SubC state is resistant to small external influences: adding one SOU on the network will not qualitatively change the network behavior.

At  $t_c$  moment of time, the number of SOUs reaches its critical value  $S_c$ , and the network goes into the SOC state. The SOC state is not resistant to small external influences: just one added SOU can cause an avalanche of microposts of any size. The behavior of avalanches of microposts distribution in a self-organized critical network is unpredictable based on the behavior of its individual users. In this case, the social network has the emergence property and in this sense is a complex system. Thus, the cause of the emergence of microposts avalanches of all sizes is the self-organized criticality of a certain network community, which users are united by interest in some topics. It should be noted that the network in the SubC state does not necessarily go to the SOC state. It is possible to relax the network until it reaches the SOC state. Relaxation may be caused by the decrease or the constancy of the SOUs number over time due to the loss of interest of SOUs in the topic discussed on the network.

Twitter self-organization in a critical state occurs when the number of microposts ( $\eta$ ) relevant to a certain topic barely becomes nonzero ( $+0$ ), i.e., corresponds to its separation from zero ( $\eta = +0$ ). To ensure  $\eta = +0$ , one SOU is added in Twitter at each time interval  $\Delta t$ , corresponding to the relaxation time. The SOC state is robust in relation to possible changes on the social network. For example, if the nature of interactions between users changes, the social network temporarily deviates from the existing critical state, but after a while, it is restored in a slightly different form. The hierarchical network structure will change, but its dynamics will remain critical. Every time, when trying to divert Twitter from the SOC state, the social network invariably returns to this state.

The regular return to the SOC state for any deviations from it let us suggest that it is a special kind of stable equilibrium of the evolving network, which, according to Bak, is called a punctuated equilibrium [7, 8]. If the social network is in such an equilibrium, then significant changes in it can occur both with a strong external impact (for example, with the strong social network pumping by strategically oriented users) and as a result of gradual internal changes.

The ordered SupC state, observed if  $S > S_c$ , is resistant to small external influences: adding one SOU on the network will not qualitatively change the network behavior. In this state, characterized by a supercritical number of SOUs, the size of the avalanches of microposts continues to grow. If the

network does not pump by SOUs, then it relaxes, returning back to the SubC state.

*2.2. Spectrum of Self-Organized Criticality Exponents.* In Section 2.1, it was noted that the presence of avalanches of large microposts on the network is a characteristic feature of being Twitter in the SOC or the SupC state. Consequently, the results of a quantitative analysis of the avalanche sizes of microposts can be used as an identifier of the social network state.

To determine the network state, it is necessary to determine the size of avalanche microposts, which will allow the social network to be assigned to one of the critical states.

Considering that the theory of self-organized criticality is one of the foundations of the complexity theory (sometimes called the paradigm) [29], we will use the more general concept of complexity: a nonstrict definition of complexity at the level of external demonstrations of criticality of the system regardless of its internal structure. In this case, the complex system is the system, which is capable of generating extremal events: unexpected (unpredictable) and/or extraordinary events.

In the case of Twitter, we are talking about certain features of the observed time series of microposts ( $\eta_t$ ), for example, the presence of sharply allocated values of the time series. Another feature is the existence of sharply increasing sequence of time series values up to a critical value corresponding to the distribution of the avalanche of microposts on the network.

The key features of the complexity of the social networks at the level of the time series generated by them are the power law for the probability distribution function (power-law PDF) of the time series of microposts, the power spectral density (PSD) of the time series, which is characterized by  $1/f$  noise, and the power law for the autocorrelation function (power-law ACF), which is characterized by the presence of the long memory in the time series [30].

In the general case, the power-law PDFs can be considered as a statistical value of the scale invariance of the time series of microposts:

$$p(\eta) \propto \eta^{-\alpha}, \quad (1)$$

where  $\alpha \in (1, 3)$ . It should be noted that usually power-law PDFs are characterized by  $\alpha \in (2, 3)$  [31]. We consider the most common case belonging to power-law PDF. Power laws with  $\alpha \leq 1$  cannot be normalized and are usually not found in natural phenomena. Power-law PDF with a low value of  $\alpha$  does not have a finite average  $\eta$  ( $\eta \rightarrow \infty$  for  $\alpha \leq 2$ ), but for  $\alpha > 2$  the average is defined. The mean square  $\eta^2 \rightarrow \infty$  for  $\alpha \leq 3$ , but  $\eta^2$  has a well-defined value for  $\alpha > 3$ .

Power-law PDF (1) refers to distributions with heavy tails, for which, unlike compact distributions, the well-known  $3\sigma$  rule (the possibility of neglecting the values of the number of microposts exceeding  $3\sigma$ ) is not satisfied. If the distribution (1) is fulfilled, then rare large events do not occur infrequently enough for their probability to be neglected. The possibility of gigantic, extraordinary events



appearing on Twitter indicates the network's tendency for disasters.

Another characteristic of the scale-invariant properties of the time series is  $1/f$  noise, which is observed in the power form of PSD at low frequencies  $f$ :

$$S(f) \propto f^{-\beta}. \quad (2)$$

The  $\beta$  value in PSD (2) determines the color of the noise. For  $1/f$  noise,  $\beta \in (0.5, 1.5)$ . The case of  $\beta = 1$  is usually referred to as pink noise.  $1/f$  noise is characteristic of all complex systems, regardless of their nature. If in the time series  $\eta_t$  there is  $1/f$  noise, then for the social network there are no periodically repeated values of the number of microposts. This is due to the fact that, in the time series of microposts, it is impossible to distinguish one characteristic scale responsible for the appearance of large values of the number of microposts. The scale-invariant type of PSD demonstrates a strong nonlinearity of social network signals when it is impossible to isolate individual components in the spectrum and offer its physical interpretation. Thus, the dynamics of Twitter microposts, in which  $1/f$  noise is observed, cannot be decomposed into separate components. Twitter, operating in a self-organized state, generates oscillations of microposts with PSD of the form (2).

The third universal feature of complexity associated with power laws (1) and (2) is the existence of the long memory in the time series of microposts. In simple systems, the time correlation function (for example, the autocorrelation function), which shows the extent of which the time series "remembers" its history, has the following form [30]:

$$\rho(\tau) \propto \exp\left(\frac{-\tau}{\tau_s}\right). \quad (3)$$

Complex systems are characterized by a power-law decrease in ACF as the time lag  $\tau$  increases:

$$\rho(\tau) \propto \tau^{-\gamma}, \quad (4)$$

where  $\gamma \in (0, 1)$ .

The existence of power-law ACF for the time series of microposts means that the current number of microposts largely depends on the past number of microposts generated by Twitter, as well as the absence of characteristic times at which information about the previous appearance of microposts would be lost. In addition, dependence (4) determines a slow power decrease in the probability of a microposts flow at time  $t$  under the condition that the same flow appeared earlier at time  $t_0$  ( $t > t_0$ ).

It is fundamentally important that the existence of long temporal correlations states the fact of the emergence of Twitter. This fact determines the possibility of the emergence of the avalanche of microposts (extremal events) as a result of the coordinated behavior of strategically oriented network users. The mechanism of occurrence of emergent Twitter properties is described in detail in Section 2.1. If for the time series of microposts relevant to a certain topic, power laws (1), (2), and (4) are fulfilled, then the following important consequences are possible.

Firstly, the relevant Twitter segment, which includes SOUs and ROUs, distributing microposts relevant to a particular topic, is in the SOC state. Secondly, power laws describe large-scale invariance in the structure of time series of microposts generated by the self-organized critical social network. The approach to the study of scale invariance is considered in Section 3.2.

PDF, PSD, and ACF in the form of power laws make it possible to use the range of interval indicators  $\alpha, \beta, \gamma$  as the indicator of the self-organized criticality (complexity) of the social network (Twitter). If the social network is in the SOC or the SupC states, then for such states the indicators of power laws take the values from the intervals  $(1, 3)$ ,  $(0.5, 1.5)$ ,  $(0, 1)$ . Otherwise, Twitter is in the SubC state.

In conclusion of this section, we note that the proposed approach to identifying the network complexity is not based on a statistical analysis of its graph structure, but on a statistical and fractal analysis of time series generated by the network.

According to the definition of Dorogovtsev and co-authors [4], "complex networks are networks with more complex architectures than classical random graphs with their 'simple' Poissonian distributions of connections." These networks are networks with heavy-tailed distributions, in particular, with the power-law distributions. One of the complex networks classes are scale-free networks. The definition of the scale-free network at the level of its graph structure was proposed by Barabási and Albert more than 20 years ago [32]. The network is scale-free, if the distribution function of the vertices  $u$  by the number of edges  $k$  is determined by a power law:

$$u(k) \propto k^{-\alpha}, \quad (5)$$

where, as well as in (1),  $\alpha \in (2, 3)$ . But usually a network is considered as scale-free if  $\alpha > 1$ . For example, there are scale-free networks with  $\alpha > 3$  [33].

It should be noted that Barabási's preferential attachment is not the only one mechanism for scale-free networks to arise; there are several other mechanisms (e.g., see the works [32, 34, 35]). For further discussions, it is important that dependency (5) is satisfied regardless of the mechanism.

There are many studies that present an empirical justification of the feasibility of equation (5) for a large number of different types of social networks (e.g., see the works [36–38]). However, recent studies have appeared showing that not for all social networks, the power law (5) is statistically justified (e.g., see the works [39, 40]). It turned out that the identification of power laws of the distribution of vertices in natural or artificial systems is not so simple (e.g., see the works [39, 41, 42]). For example, it is not always possible to distinguish a power law from a lognormal one for samples of small size. A parabola corresponding to a lognormal law in logarithmic coordinates on a sufficiently small interval of values  $k$  looks like a straight line corresponding to a power law.

What is the advantage of our proposed approach?

First of all, there is no need to divide scale-free networks into several types depending on the value of the assessment of the indicator  $\alpha$  and the level of significance with which this assessment was done, as proposed in [42]. Indeed, our approach involves the use of the three network complexity indicators  $\alpha, \beta, \gamma$ . Using the spectrum of indicators of complexity  $\alpha, \beta, \gamma$  as a network complexity identifier is only possible in the case of analysis of the time series generated by the network (realizations of some random process), but not when analyzing the data of the static network structure. Indeed, PSD and ACF are characteristics of signals, random processes, and time series.

Secondly, the use of the spectrum  $\alpha, \beta, \gamma$  to identify network complexity has a rationale within the paradigm of the complexity, as one of the paradigms of nonlinear science, and, being its core, the theory of self-organized criticality. In the context of this theory, the network complexity is determined only by the values of the spectrum indicators and does not depend on the distribution type  $u(k)$ .

Third, the use of the spectrum  $\alpha, \beta, \gamma$  allows us to identify the subcritical, the self-organized critical, and the supercritical states of the network operation.

Fourth,  $\beta$  and  $\gamma$  indicators can be independently estimated both as a static estimate of the slope ratio in a log-log scale and as a result of an estimation of the scaling indicator of time series of microposts, for example, using detrended fluctuation analysis (see Section 3.2).

### 3. Data Analysis, Results, and Discussion

This section provides a brief overview of using data mining techniques which are necessary for the formation of the time series of microposts and their statistical and fractal analysis, as well as the evaluation of Twitter complexity indicators and their interpretation.

**3.1. Mining Twitter Time Series Data.** The most suitable data source for mining of Twitter time series data that contain tweet ids (unique identifiers of tweets) regarding different events, such as political elections and natural disasters, is Harvard Dataverse. It contains the datasets of tweets ids on 12 different topics, and each dataset consists of more than 2 million unique tweet ids in the form of the 18-digit numbers (for example, 1128408193699340294) combined into one text file (.txt). Harvard Dataverse collected data using Social Feed Manager, which is the open source software that harvests social media data and web resources from Twitter. The reason why it is necessary to start to work with tweets ids, rather than tweets itself, is the fact that per Twitter's Developer Policy, tweet ids may be publicly shared for academic purposes, but tweets may not.

Nevertheless, in order to get Twitter time series, it is necessary to hydrate the obtained datasets of tweet ids. Hydrating is the process of loading JSON objects from tweets based on available tweet ids. It can be done using the

API-interface of Twitter, as well as using third-party applications. We did it with a Hydrator version 0.0.3 software. According to the obtained data, it is possible to build the interaction structure of users and time series of tweets (including retweets and other mentions).

We used the following relevant tweet ids time series events and themes for the formation and subsequent statistical and fractal analysis of the time series of microposts:

- (1) 2016 United States Presidential Election Tweet Ids [43]. The dataset contains the tweet ids of approximately 280 million tweets and retweets related to the 2016 United States Presidential Election. Tweets were collected between July 13, 2016, and November 10, 2016.
- (2) Women's March Tweet Ids [44]. The dataset consists of the tweet ids of 7,275,228 tweets and retweets related to the Women's March on January 21, 2017. Tweets were collected between December 19, 2016, and January 23, 2017.
- (3) End of Term 2016 US Government Twitter Archive [45]. The dataset consists of the tweet ids of 5,655,632 tweets and retweets, and the original tweets were made from approximately 3000 Twitter accounts, which are connected with the US government. Tweets were collected between October 21, 2016, and January 21, 2017.
- (4) Hurricanes Harvey and Irma Tweet Ids [46]. The dataset consists of the tweet ids of 35,596,281 tweets and retweets related to Hurricanes Irma and Harvey.
- (5) Immigration and Travel Ban Tweet Ids [47]. The dataset consists of the tweet ids of 16,875,766 tweets and retweets related to the immigration and travel ban that was announced by the Trump Administration in January 2017. Tweets were collected between January 30, 2017, and April 20, 2017.
- (6) Charlottesville Tweet Ids [48]. The dataset consists of the tweet ids of 7,665,497 tweets and retweets related to events in Charlottesville, Virginia, in August 2017.
- (7) Winter Olympics 2018 Tweet Ids [49]. The dataset consists of the tweet ids of 13,816,206 tweets and retweets related to the 2018 Winter Olympics held in Pyeongchang, South Korea. Tweets were collected between January 31, 2018, and February 27, 2018.
- (8) US Government Tweet Ids [50]. The dataset consists of the tweet ids of 9,673,959 tweets and retweets, and the original tweets were made from approximately 3400 US government accounts. These accounts are linked with the federal US government agencies. Tweets were collected between January 20, 2017, and July 20, 2018.
- (9) News Outlet Tweet Ids [51]. The dataset consists of the tweet ids of 39,695,156 tweets and retweets, and the original tweets were made from the Twitter

accounts of approximately 4500 news outlets; it means accounts of mass media that intended to disseminate news. Tweets were collected between August 4, 2016, and July 20, 2018.

- (10) 2018 US Congressional Election Tweet Ids [52]. The dataset consists of the tweet ids of 171,248,476 tweets and retweets related to the 2018 US Congressional Election. Tweets were collected between January 22, 2018, and January 3, 2019.
- (11) 115th US Congress Tweet Ids [53]. The dataset consists of the tweet ids of 2,041,399 tweets and retweets, and the original tweets were made from the Twitter accounts of members of the 115th US Congress. Tweets were collected between January 27, 2017, and January 2, 2019.
- (12) Ireland 8th Tweet Ids [54]. The dataset consists of the tweet ids of 2,279,396 tweets and retweets related to the referendum to repeal the 8th amendment to the Irish constitution on May 25, 2018. Tweets were collected between April 13, 2018, and June 4, 2018.

As a result, we got twelve equidistant (a step is 1 second) time series of microposts  $\{\eta_i\}$  ( $i = 1, 2, \dots, N$ ) of different lengths  $N$ , each of which is relevant to some topic (tweet Ids). Next, there is a description of the time series analysis methods  $\{\eta_i\}$  (see Section 3.2), as well as the results of such analysis for each time series, obtained over its entire length  $N$  (see Section 3.3).

**3.2. Statistical and Fractal Methods for Twitter Time Series Analysis.** In the context of our study, the main purpose of analysis of the time series of microposts is to statistically confirm the statement that the range of empirical indicators of complexity  $\alpha, \beta, \gamma$  takes values  $(1, 3), (0.5, 1.5), (0, 1)$ . Formally, we tested the statistical hypothesis of the significance of a simple linear regression:

$$\ln y_i = \delta_0 + \delta_1 \ln x_i + \varepsilon_i, \quad (6)$$

where  $y_i \equiv p_i$ ,  $x_i \equiv \eta_i$ ,  $\delta_1 \equiv \alpha$  for PDF,  $y_i \equiv S_i$ ,  $x_i \equiv f_i$ ,  $\delta_1 \equiv \beta$  for PSD, and  $y_i \equiv \rho_i$ ,  $x_i \equiv \tau_i$ ,  $\delta_1 \equiv \gamma$  for ACF. This is a test of the null hypothesis  $H_0: \delta_1 = 0$ , with an alternative hypothesis  $H_1: \delta_1 \neq 0$ . As a measure of agreement with the null hypothesis,  $p$  value was used as an indicator of the minimum level of significance, in which  $H_0$  being rejected. We used the ordinary least squares (OLS) method for estimating the parameters  $\delta_0$  and  $\delta_1$ .

The possibility of transition to simple linear regressions for statistical analysis of the time series is due to the scale invariance of the dependences (1), (2), and (4). On the one hand, such a transition will make it easier to obtain statistical estimates of the indicators  $\alpha, \beta$ , and  $\gamma$ ; on the other hand, it will allow to establish a filter for separating power laws from other non-scale-invariant laws, for example, from normal, exponential, lognormal, and extended exponential laws for PDF.

The ACF for the observed time series  $\{\eta_i\}$  represents the correlation of the values  $\eta_i$  and  $\eta_{i+\tau}$  for different time lags  $\tau$

( $\tau = 0, 1, 2, \dots, N$ ), i.e., correlations over different time scales  $\tau$ . It is calculated based on formula  $\rho(\tau) = \langle \Delta\eta_i \Delta\eta_{i+\tau} / \Delta\eta_i^2 \rangle$ , where  $\Delta\eta_i = \eta_i - \eta_{i-1}$  is an increment and  $\langle \Delta\eta_i \Delta\eta_{i+\tau} \rangle$  and  $\langle \Delta\eta_i^2 \rangle$  state the mean value for the data  $\{\Delta\eta_i \Delta\eta_{i+\tau}\}$  and  $\{\Delta\eta_i^2\}$ , respectively. We applied standard spectral analysis techniques (Fourier transform) to calculate the PSD  $S(f)$  of the time series  $\{\eta_i\}$  as a function of the frequency  $f$ .

The traditional approach to the time series analysis relies on the measurement of PSD and ACF. However, only the implementation of Gaussian processes is exhaustively described by their second moments. Outside of such implementations, a complete statistical description requires an estimate of higher order moments. In addition, higher order moments do not always have such a clear physical meaning as ACF and PSD. Therefore, evaluations of a small number of values that can be given a certain meaning become important. These values include the fractal dimensions of the time series.

The fractal dimension is closely related to the scaling index  $s$ , which can be the Hurst exponent, estimated by the method of normalized range or fluctuation analysis (FA) [55], or the generalized Hurst exponent, estimated by the method of detrended fluctuation analysis (DFA) [56].

The DFA method is an efficient method for analysis of the time series characterized by the presence of the long memory or  $1/f$  noise. The DFA method is a generalization of the FA method for analysis of the scale invariance of nonstationary time series.

The DFA method allows both to estimate the scaling indicator of the time series  $s$  and to obtain indirect estimates of  $\beta_s$  and  $\gamma_s$  indicators, calculated from the generalized scaling indicator  $s$  of the time series. In the first case, it is about the definition of  $s$  of scale-invariant (in the narrow sense) time series  $\eta_t$ ,  $t \in \mathbb{N}$ , i.e., time series for which the equality of probability distributions ( $\eta_{at}$  and  $a^s x_t$ ) takes place [57]; in the second case, it is about the existence of dependencies for the time series with power-law ACF and  $1/f$  noise of the form [58]:

$$\begin{aligned} \gamma_s &= 2 - 2s, \\ \beta_s &= 2s - 1. \end{aligned} \quad (7)$$

The FA method does not always give correct estimates of the indicator  $s$  for the most time series [30]. Compared with the FA method, the DFA method gives more correct estimates in most cases [30], so we used this method to estimate  $s$ .

The DFA method is one of the algorithms based on the ideology of the transition from the original time series  $\eta_t$ ,  $t \in \mathbb{N}$  to the generalized model of one-dimensional random walks. In this algorithm, the data are first reduced to a null average with the subsequent construction of a random walk  $x_k = \sum_{i=1}^k |\eta_i - \bar{\eta}|$ . Next, the series  $x_k$ ,  $k \in \mathbb{N}$  is divided into nonintersecting segments of length  $a$ , within each of which the equation of a straight line is defined, approximating the sequence  $x_k$ . Found  $x_k(a)$  approximation is treated as a local trend. Next, the mean square error of the linear approximation is calculated:



$$F(a) = \left[ \left( \frac{1}{N} \right) \sum_{k=1}^N (x_k - x_k(a))^2 \right]^{0.5}, \quad (8)$$

and the corresponding calculations are made in a wide range of values  $a$ . It is believed that the dependence  $F(a)$  often has the power form  $F(a) \propto a^s$ , and the presence of a linear segment in a log-log scale allows us to state that scaling exists.

Numerical values characterize different types of correlated dynamics of microposts, if  $s \neq 0.5$  and uncorrelated behavior at  $s = 0.5$ . For example, the interval  $0 < s < 0.5$  corresponds to anticorrelations (the alternation of large and small values in the time series of microposts);  $0.5 < s < 1$  determines the correlated dynamics (large compared to the average values more often follow large values, and small values follow small ones). The special case  $s = 1$  is observed for  $1/f$  noise.

**3.3. Data Analysis Results.** Table 1 presents the OLS estimates of the spectrum complexity as a slope  $\delta_1$  of linearized dependencies (6) and DFA estimates of scaling indicators  $s$ ,  $\beta_s$ , and  $\gamma_s$ , obtained using the dependencies (7). The corresponding  $p$  values are shown in brackets.

The symbol “—” denotes the absence of statistically significant DFA estimates for  $\beta_s$  and  $\gamma_s$  indicators (see equation (7)). This is due to the fact that there are no statistically significant linear dependencies of  $\ln F(a) \propto s \ln a$  for the corresponding time series of microposts. Statistically significant values of the exponents are denoted in bold.

**3.4. Results and Their Discussion.** The most significant result in the context of our study is the existence of two classes of time series of microposts and tweet Ids corresponding to them.

The first class consists of time series for which  $\alpha \in [1.23, 2.23]$ ,  $\beta \in [1.05, 1.29]$ , and  $\gamma \in [0.12, 0.43]$ . Indicators of the power laws of such time series belong to the spectrum of indicators of complexity (1, 3), (0.5, 1.5), (0, 1) and, consequently, Twitter, which generates such time series of microposts, is in the SOC state or the SupC state. The social network is capable of generating extreme events, which are avalanches of microposts of all sizes (regarding “sizes,” see Section 2.1) corresponding to the following tweet ids: “2016 United States Presidential Election,” “Women’s March,” “Hurricanes Harvey,” “Hurricanes Irma,” “Immigration and Travel Ban,” “Charlottesville,” “2018 US Congressional Election,” and “Ireland 8th.” In addition, the current number of microposts largely depends on the past number of microposts generated by Twitter. Indeed, for all the time series of this class indicator ACF,  $\gamma \in (0, 1)$ . It is noteworthy that all tweet ids relate either to protest movements or to political elections or to the population activities during natural disasters. PDF of such time series have infinite  $\eta^2$  and infinite  $\eta$  for events related to political elections and finite  $\eta$  in all other cases. DFA estimates of  $\beta_s$  and  $\gamma_s$  give close values to the corresponding indicators  $\beta$  and  $\gamma$ , and the presence of statistically significant values of

the scaling exponent  $s$  determines the scale invariance of time series, which is one of the key features of the self-organized criticality of the social network. In addition, for all time series of the first class  $s \equiv 1$  and  $\beta_s \equiv 1$ , which corresponds to the presence of pink noise and, accordingly, being Twitter in the SOC or the SupC states. The existence of a dependency (4) for the time series of microposts means that the current numbers of microposts largely depend on the past number of microposts generated by Twitter, as well as the absence of characteristic times at which information about previous occurrences of microposts would be lost.

The second class consists of time series for which  $\alpha \in [3.24, 3.99]$ ,  $\beta \in [0.19, 0.26]$ , and  $\gamma \in [5.24, 6.01]$ ; moreover, estimates of all indicators are not statistically significant: statistical hypothesis is accepted with previously considered  $p$  values shown in Table 1. Consequently, for these time series of microposts, at least the power laws (1) and (4) are not satisfied. This result is consistent with the results of the detrended fluctuation analysis, according to which there is no statistically significant estimate of the scaling exponent  $s$ ; therefore, these time series of microposts are not scale-invariant. Thus, Twitter, which generates these time series, is neither in the SOC state nor in the SupC state. Twitter users, that is, in such a state, are not coordinated. This leads to the generation of the time series, for which the spectrum is not performed. It may be the SubC state, but such a conclusion requires the determination of the explicit form of PDF and ACF dependencies, which is beyond the scope of our study. The only argument in favor of the assumption of Twitter being in the SubC state is the fact that indicator  $\beta \in [0.19, 0.26]$  is close to the value corresponding to white noise ( $\beta \equiv 0$ ).

## 4. Three-Parameter Twitter Self-Organization Model

The results of the analysis of the time series of microposts presented in Section 3 are important not only for solving management problems and identifying the state of Twitter, but also as the basis for the development and verification of macroscopic models describing evolutionary processes on the social network. In the context of our study, the analysis of such time series is necessary for the verification of the model, describing the SOC state of Twitter. The construction and verification of such model is the purpose of the research presented in Section 4.

A sufficient verification of Twitter’s self-organized criticality model is that the indicators of the power laws (1), (2), and (4) of the theoretical and observable time series of microposts belong to the spectrum of complexity indicators (1, 3), (0.5, 1.5), (0, 1).

**4.1. Generalized Three-Parameter Model.** It is known (e.g., see the works [59–61]) that the concept of self-organization is a generalization of the physical concept of critical phenomena, such as phase transitions. Therefore, the phenomenological theory that we propose is a generalization of the theory of thermodynamic transformations for open



TABLE 1: Estimates of the values of the complexity spectrum indices for time series of microposts.

Tweet Ids	$\alpha$	$\beta$	$\gamma$	$\beta_s$	$\gamma_s$	$s$
2016 United States Presidential Election	<b>1.23</b> (0.0121)	<b>1.29</b> (0.0182)	<b>0.12</b> (0.0201)	<b>0.92</b> (0.0036)	<b>0.08</b> (0.0036)	<b>1.04</b> (0.0036)
Women's March	<b>2.11</b> (0.0234)	<b>1.23</b> (0.0198)	<b>0.42</b> (0.0211)	<b>0.90</b> (0.0101)	<b>0.10</b> (0.0101)	<b>1.05</b> (0.0101)
End of Term 2016 US government	3.24 (0.6743)	0.24 (0.7235)	5.24 (0.6990)	—	—	0.45 (0.7699)
Hurricanes Harvey	<b>2.12</b> (0.0312)	<b>1.13</b> (0.0289)	<b>0.34</b> (0.0320)	<b>0.89</b> (0.0015)	<b>0.11</b> (0.0015)	<b>1.06</b> (0.0015)
Hurricanes Irma	<b>2.23</b> (0.0234)	<b>0.98</b> (0.0194)	<b>0.18</b> (0.0209)	<b>0.96</b> (0.0098)	<b>0.04</b> (0.0098)	<b>1.02</b> (0.0098)
Immigration and Travel Ban	<b>2.18</b> (0.0401)	<b>1.09</b> (0.0320)	<b>0.21</b> (0.0128)	<b>0.97</b> (0.0094)	<b>0.03</b> (0.0094)	<b>1.02</b> (0.0094)
Charlottesville	<b>2.18</b> (0.0313)	<b>1.21</b> (0.0287)	<b>0.43</b> (0.0121)	<b>0.90</b> (0.0101)	<b>0.10</b> (0.0101)	<b>1.05</b> (0.0101)
Winter Olympics 2018	3.59 (0.7239)	0.22 (0.6348)	5.64 (0.5341)	—	—	0.52 (0.8172)
US Government	3.28 (0.6361)	0.19 (0.7298)	6.01 (0.6399)	—	—	0.48 (0.7456)
News Outlet	3.36 (0.4275)	0.23 (0.3895)	5.50 (0.4458)	—	—	0.54 (0.6451)
2018 US Congressional Election	<b>1.47</b> (0.0281)	<b>1.05</b> (0.0398)	<b>0.22</b> (0.0435)	<b>0.95</b> (0.0099)	<b>0.05</b> (0.0099)	<b>1.03</b> (0.0099)
115th US Congress	3.99 (0.3189)	0.26 (0.4197)	5.24 (0.5618)	—	—	0.46 (0.9999)
Ireland 8 <sup>th</sup>	<b>2.18</b> (0.0311)	<b>1.18</b> (0.0270)	<b>0.35</b> (0.0311)	<b>0.97</b> (0.0129)	<b>0.03</b> (0.0129)	<b>1.02</b> (0.0129)

systems. Twitter self-organization is possible due to its openness, since there are incoming and outgoing network flows of its users constantly; its macroscopic nature, because it includes a large number of users; and its dissipation, because there are losses in the flows of microposts and associated information.

Based on the synergetic principle of subordination, it can be argued that Twitter's self-organization in a critical state is completely determined by the suppression of the behavior of an infinite number of microscopic degrees of freedom by a small number of macroscopic degrees of freedom. As a result, the collective behavior of users of the social network is defined by several parameters or degrees of freedom: an order parameter  $\eta_t$ , its role is the number of microposts relevant to a certain topic that are sent by SOUs and, unwittingly following their strategies, by ROUs; a conjugate field  $h_t$  is information associated with microposts distributed on the network; and a control parameter  $S_t$  which is the number of SOUs of the networks. On the other hand, in Twitter's self-organization as the nonequilibrium system, the dissipation of flows of microposts on the network should play a crucial role, which ensures the transition of the network to the stationary state. In the process of self-organization in a critical state of the network, all three degrees of freedom have an equal character, and the description of the process requires a self-consistent view of their evolution. The restriction to three degrees of freedom is also determined by the Ruelle–Takens theorem [62], according to which a nontrivial picture of self-organization is observed if the number of selected degrees of freedom is, at least, three.

Kinetic equations and a detailed physical substantiation of the relations between its parameters are given in our paper [63]. The construction of the three-parameter self-organization scheme was based on the analogy between the mechanisms of functioning of a single-mode laser and the microblogging social network. The study of possible modifications of equations leading to models that are capable of describing critical phenomena on Twitter, in particular the SOC or the SupC states, is outside of the scope of this paper. These equations in dimensionless quantities have the following form:

$$\begin{cases} \dot{\eta}_t = -\eta_t^\varepsilon + h_t + \sqrt{I_\eta} \xi_t, \\ \frac{\tau_h}{\tau_\eta} \dot{h}_t = -h_t + \eta_t^\varepsilon S_t + \sqrt{I_h} \xi_t, \\ \frac{\tau_S}{\tau_\eta} \dot{S}_t = (S_0 - S_t) - \eta_t^\varepsilon h_t + \sqrt{I_S} \xi_t, \end{cases} \quad (9)$$

where  $I_i$  is the intensity of fluctuations (or noises) of each of the degrees of freedom  $i = \eta, h, S$ ;  $\tau_i$  is relaxation times of corresponding quantities;  $\xi_t$  is the white noise due to random factors; and  $S_0$  is the number of SOUs of the network at the initial time moment ( $t = 0$ ) of the network evolution. The parameter  $S_0$  determines the degree of external disturbance or pumping of the social network by strategically oriented users, which removes Twitter from the equilibrium state.

Assuming  $\varepsilon = 1$ , and also neglecting random factors  $\xi_t$ , equation (9) represents a well-known system of Lorenz equations, in which dynamic variables describe the self-consistent behavior of the order parameter, the conjugate field, and the control parameter. In such a system, the functions  $\eta_t/\tau_\eta$ ,  $h_t/\tau_h$ , and  $(S_t - S_0)/\tau_S$  describe autonomous relaxation of the number of microposts, of conjugate information, and the number of strategically oriented network users to stationary values  $\eta_t = 0$ ,  $h_t = 0$ ,  $S_t = S_0$ . The Lorenz system takes into account Le Chatelier's principle: since the reason for self-organization is the growth of the control parameter  $S_t$ , the values of  $\eta_t$  and  $h_t$  should be changed in such a way as to prevent the growth  $S_t$ . Finally, the positive feedback between the order parameter  $\eta_t$  and the control parameter  $S_t$ , which leads to an increase in the conjugate field  $h_t$ , is fundamentally important. This is what causes the self-organization [60]. Despite the fact that the Lorenz system is a rather rough approximation in solving some problems, the system is an adequate model that qualitatively describes the processes of self-organization in systems of various nature, including the kinetics of first- and second-order phase transitions [59].

The feedback intensity indicator  $\varepsilon$  in equation (9), which also distinguishes it from the Lorenz system, is an indicator of the disturbance of Twitter's ordering on its self-consistent

behavior. From a physical point of view, replacing the order parameter normalized to one ( $\eta_t \in (0, 1]$ ) with a larger value  $\eta_t^\varepsilon$  ( $\varepsilon < 1$ ) means that the ordering process affects Twitter's self-consistent behavior more than in the ideal case, when  $\varepsilon = 1$ . In the case when  $\eta_t \in (0, \infty)$ , the parameter  $\varepsilon$  can be determined by introducing the unit step function  $\theta(\eta_t)$  using the following replacement:

$$\varepsilon \equiv \varepsilon + \theta(\eta_t - 1), \quad (10)$$

$$\text{where } \theta(\eta_t - 1) = \begin{cases} 0, & \eta_t < 1, \\ 1, & \eta_t \geq 1. \end{cases}$$

Another replacement is the replacement in equation (9) of the order parameter  $\eta_t$  of the following form:

$$\eta_t^\varepsilon \equiv \text{Sign}(\eta_t) |\eta_t|^\varepsilon. \quad (11)$$

The meaning of the transition  $\eta_t \rightarrow \text{Sign}(\eta_t) |\eta_t|^\varepsilon$  in the Lorenz system is that the transition of the order parameter to its absolute value avoids the minimum values of the power function  $\eta_t^\varepsilon$  with a fractional exponent  $\varepsilon$ .

Further, if it is not specified separately, the parameters  $\varepsilon$  and  $\eta_t$  are defined as (10) and (11).

**4.2. Self-Organized Criticality of Twitter in Adiabatic Approximation.** Equation (9), as well as the system of Lorenz equations, does not have an exact analytical solution. When certain conditions are met, system (9) in an adiabatic approximation can be quite acceptable approximation. The adiabatic self-organization mode corresponds to a phase transition process for which the stationary value of the control parameter does not reduce to the pump parameter (e.g., see the works [59–61]).

In the adiabatic approximation, the characteristic relaxation time of the number of microposts  $\tau_\eta$  far exceeds the corresponding relaxation times of the information associated with microposts and the number of strategically oriented users:  $\tau_h$  and  $\tau_S$ . This means that the information  $h_t \equiv h(\eta_t)$  and the number of strategically oriented users  $S_t \equiv S(\eta_t)$  follow the changes in the number of microposts  $\eta_t$  on Twitter. When the conditions  $\tau_\eta \gg \tau_h, \tau_S$  are fulfilled, the principle of subordination makes it possible to neglect the fluctuations of the quantities  $h(\eta_t)$  and  $S(\eta_t)$  in equation (9), i.e., assume  $(\tau_h/\tau_\eta)\dot{h}_t = (\tau_S/\tau_\eta)\dot{S}_t = 0$ .

The use of the adiabatic approach to Twitter as an open nonequilibrium system means that, when the value of the social network pumping by strategically oriented users tends to zero ( $S_0 \rightarrow 0$ ), there is a slow decrease in the flow of microposts  $\eta_t$  and a rapid decrease in the associated information  $h_t$  as well as in the number of strategically oriented users  $S_t$ , who sending microposts.

Using the adiabatic approximation allows to reduce the dimension of the phase space, i.e., transit from the analysis of a three-dimensional dynamic system with additive noise (9) to the analysis of a one-parameter stochastic system with multiplicative noise:

$$\tau_\eta \dot{\eta}_t = f_\eta + \sqrt{I_\eta} \xi_t. \quad (12)$$

In the Langevin equation (12), the drift and diffusion parts are determined by the following values:

$$\begin{aligned} f_\eta &\equiv -\eta_t^\varepsilon + S_0 \eta_t^\varepsilon \mu_\eta, \\ I_\eta &\equiv I_\eta + (I_h + I_S \eta_t^{2\varepsilon}) \mu_\eta^2, \end{aligned} \quad (13)$$

where  $\mu_\eta \equiv 1 + \eta_t^{2\varepsilon}$ .

#### 4.3. Results and Their Discussion

**4.3.1. Self-Organized Critical State of Twitter.** Suppose that the social network Twitter is self-organized into a critical state as a result of the agreed actions of SOUs and ROUs. Such a saturated network state by strategically oriented users and information is characterized by the following features: firstly, by the significant intensity of stochastic interactions between strategically oriented users ( $I_S \gg I_\eta, I_h$ ), and secondly, by the significant impact of the streamlining process on Twitter's self-consistent behavior. In this case, equality (10) holds for the feedback intensity indicator  $\varepsilon$ . In this state, even a negligible external disturbance ( $S_0 \equiv 0$ ) is enough to spread avalanches of microposts on the social network.

Therefore, equation (12), describing the being of the social network in the SOC state, will be in the following form:

$$\tau_\eta \dot{\eta}_t = -\eta_t^\varepsilon + \sqrt{I_S} \eta_t^\varepsilon \mu_\eta \xi_t. \quad (14)$$

Suppose that the homogeneous process (14) occurs on the interval  $(\eta_{\min}, 1)$  and  $\varepsilon < 1$ , where  $\eta_{\min}$  is the minimum number of microposts for which power-law PDF is performed. Then, a nonnormalized solution of the corresponding stationary Fokker–Planck equation with reflecting boundaries is given by

$$p_s(\eta) \propto I_S^{-1} \eta^{-2\varepsilon} \mu_\eta^{-2} \exp \left[ -I_S^{-1} \int_{\eta_{\min}}^{\eta} \eta'^{-\varepsilon} \mu_{\eta'}^{-2} d\eta' \right]. \quad (15)$$

The integral in PDF (15) has the following form, rather bulky for analysis:

$$\left\{ \frac{\eta^{1-\varepsilon}}{2(\varepsilon-1)\mu_{\eta'}} \left[ (3\varepsilon-1)\mu_{\eta'}^2 F_1 + \eta'^{2\varepsilon} - \varepsilon(3\eta'^{2\varepsilon} + 2) \right] \right\} \Big|_{\eta_{\min}}^{\eta}, \quad (16)$$

where  ${}_2F_1(1, (\varepsilon-1/2\varepsilon); (3\varepsilon-1/2\varepsilon); -\eta'^{-2\varepsilon})$  is the hypergeometric function.

The graph of the unnormalized PDF (15) in a log-log scale for  $\eta \in (\eta_{\min}, 1)$ ,  $\eta_{\min} = 0.001$  and the values of  $\varepsilon \in (0, 1)$  are presented in Figure 1.

Distribution (15) is a power-law PDF of the form (1) with indicator  $\alpha$ , corresponding to indicators of feedback intensity  $\varepsilon$ . Increasing the feedback intensity leads to an increase in the stationary probability  $p_s(\eta)$  for all  $\eta \in (0.001, 1)$ .

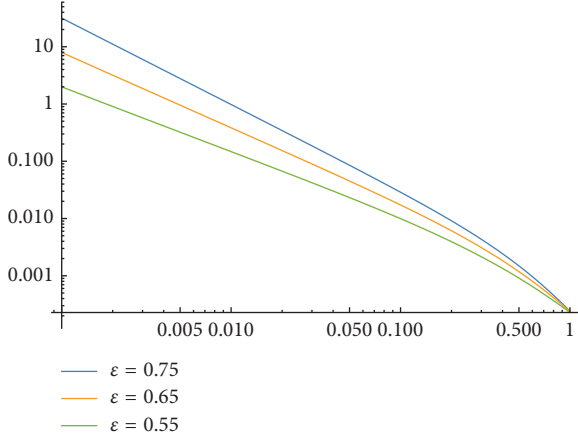


FIGURE 1: Log-log plot of the distribution (15) for  $I_S = 1000$ , and three different  $\varepsilon = 0.55, 0.65, 0.75$ .

In order to obtain analytical values for PSD and ACF of a random process (14), it is necessary to obtain an analytical solution  $p(\eta_t, t)$  of the corresponding nonstationary Fokker–Planck equation. If it is possible to obtain an exact analytic values for  $p(\eta_t, t)$  and, accordingly, an analytical definition of  $S(f)$  and  $\rho(\tau)$ , then it will be difficult to interpret. Therefore, we obtained these dependences as a result of the numerical integration of equation (14). According to the obtained realizations of this random process, the dependences  $S(f)$  and  $\rho(\tau)$  were determined.

In Figures 2 and 3, the PSD and ACF are presented in a log-log scale for one of the process implementations (14). To obtain implementations of the stochastic processes and to process the results, standard Wolfram Mathematica version 11.2 algorithms were used. First, data were generated using the “ItoProcess” function. Further, based on the data obtained using the “PowerSpectralDensity” functions, PSDs similar to those shown in Figure 2 were built. Using the “CorrelationFunction,” ACFs were found similar to those shown in Figure 3. Also in the figures are shown the following: linear approximations and the estimates of the scaling indicators  $\beta$  and  $\gamma$  which are obtained by averaging over 10,000 realizations of the OLS estimates  $\beta$  and  $\gamma$  of the random process. The straight line shown in Figure 2 is a linear approximation of the PSD with the OLS estimate of the index  $\beta = 1.1(0.009)$  that corresponds to a random process with pink noise; straight line in Figure 3 is a linear approximation of ACF with the estimate of the index  $\gamma = 0.8(0.010)$  that corresponds to a random process with the long memory. The corresponding  $p$  values are shown in brackets.

Consequently, equation (14) is a good approximation for describing the self-organized critical state of Twitter. Indeed, all theoretical indicators of complexity (criticality)  $\varepsilon, \beta$ , and  $\gamma$  belong to the spectrum of indicators of complexity  $(1, 3), (0.5, 1.5), (0, 1)$ .

**4.3.2. Subcritical and Supercritical States of Twitter.** The SubC state of Twitter is a chaotic state characterized by the

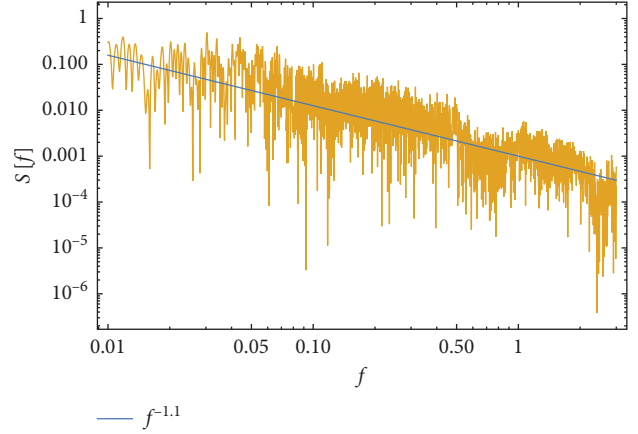


FIGURE 2: Log-log plot of the PSD for  $\tau_\eta = 1$ ,  $\varepsilon = 0.75$ , and  $I_S = 1000$ .

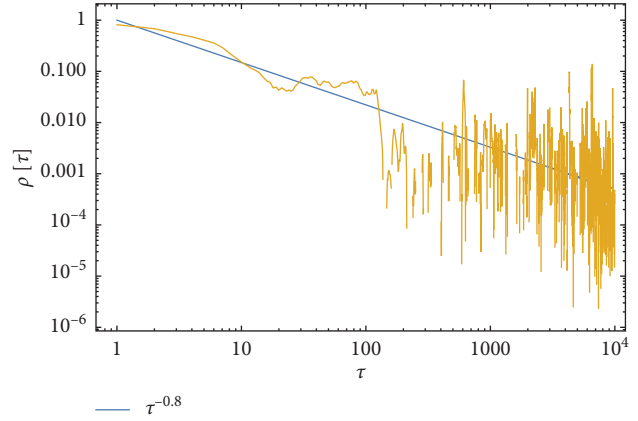


FIGURE 3: Log-log plot of the ACF for  $\tau_\eta = 1$ ,  $\varepsilon = 0.75$ , and  $I_S = 1000$ .

presence of a negligible number of avalanches of micro-posts and, therefore,  $p_s(\eta)$  is not the power-law PDF. Also SubC state is characterized by resistance to small disturbances. In this state, minor chaotic directed flows of microposts are created by all users of the social network, regardless of the size of its pumping by strategically oriented users. The social network functions in the SubC state until  $S_t$  reaches a certain critical value  $S_c$ . In this state, the streamlining process has almost no effect on the self-consistent behavior of the network, i.e., the feedback intensity indicator  $\varepsilon = 1$ , and the fluctuation intensity of each of the degrees of freedom is comparable (let us take  $I_\eta = I_h = I_S = I$ ).

Therefore, the SubC state of Twitter is described by the following Langevin equation:

$$\tau_\eta \dot{\eta}_t = -\eta_t + S_0 \eta_t \mu_\eta + \sqrt{I(1 + \mu_\eta^3)} \xi_t, \quad (17)$$

where  $\mu_\eta \equiv 1 + \eta_t^2$ .

Assuming that the process (17) is on the interval  $(0, 1)$ , the solution of the corresponding stationary Fokker–Planck equation is a PDF of the following form:

$$p_s(\eta) \propto I^{-1} (1 + \mu_\eta^3)^{-1} \exp \left[ I^{-1} \int_0^\eta \frac{S_0 \eta' \mu_{\eta'}^2 - \eta'}{1 + \mu_{\eta'}^3} d\eta' \right]. \quad (18)$$

The integral in the distribution (18) has the following form:

$$\left\{ \frac{1}{12I} \left[ 2\sqrt{3} (S_0 - 1) \tan^{-1} \left( \frac{2\eta'^2 + 1}{\sqrt{3}} \right) - (S_0 + 1) \ln \frac{(2 + \eta'^2)^{1/2}}{\eta'^4 + \eta'^2 + 1} \right] \right\} \Big|_0^\eta. \quad (19)$$

Distribution graph (18) is presented in Figure 4 in a log-log scale.

It is obvious that PDF presented in Figure 4 is not a power-law PDF and, therefore, such a distribution describes the appearance of avalanches of microposts only of relatively small size. In addition, an increase in the network pumping  $S_0$  leads to an increase, although not significant, in the frequency of appearance of relatively large numbers of microposts.

Thus, it is reasonable to assume that a further increase in the number of strategically oriented social network users to a certain critical value  $S_c$ , accompanied by a significant increase in the intensity of stochastic interactions between them ( $I_S \gg I_\eta, I_h$ ), will lead the network to self-organization in a critical state.

Twitter in the SubC state is able to generate time series of microposts with relatively small values. Perhaps, these time series correspond to the following tweet identifiers: “End of Term 2016 US Government,” “Winter Olympics 2018,” “US Government,” and “News Outlet.”

If  $S > S_c$ , then chaos changes in order, and instead of insignificant chaotic directed flows of microposts, a dedicated directional flow (avalanche) of microposts appears on the network. This flow becomes significant at the macrolevel. Like the SubC state, the SupC state is resistant to small disturbances (for example, adding only one SOU). In the SupC state, small disturbances cannot tangibly affect the size of the avalanche of microposts.

The distribution of the number of microposts which is characterized by the SupC state of Twitter is presented in Figure 5 in a log-log scale. PDFs are presented (see distribution (15)) for  $\varepsilon = 0.75$  and various values of the network pumping parameter.

The distributions shown in this figure correspond to the power-law PDF. Moreover, the weighting of the distribution tails is due to the increased pumping of the network by the strategically oriented users. If Twitter is in the SupC state, then the number of SOUs and, accordingly, the microposts avalanche sizes continue to grow.

## 5. Discussion

The obtained results are of interest both for identifying the SubC state or the SupC state of Twitter that are stable to small disturbances based on the analysis of the observed time series of microposts and for determining the causes of the social network self-organization in a critical state.

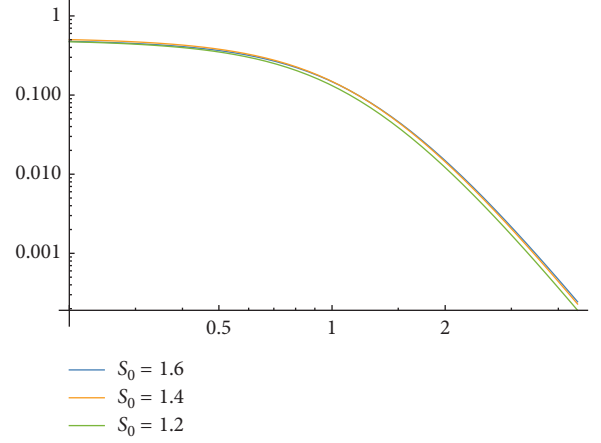


FIGURE 4: Log-log plot of the distribution (18) for  $I = 1$  and three different  $S_0 = 1.2, 1.4, 1.6$ .

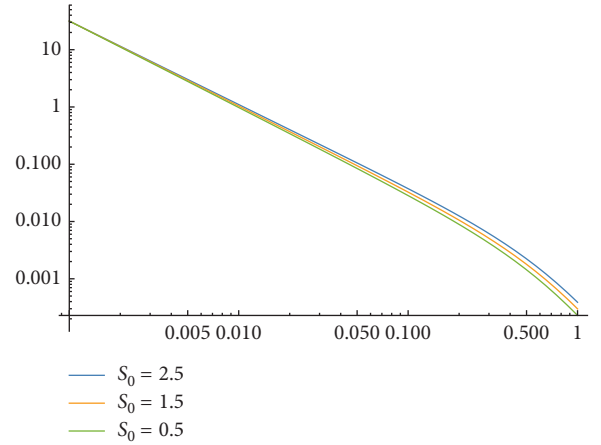


FIGURE 5: Log-log plot of the distribution (15) for  $I_s = 1000$ ,  $\varepsilon = 0.75$ , and three different  $S_0 = 0.5, 1.5, 2.5$ .

The presence of a spectrum of criticality indicators  $(1, 3), (0.5, 1.5), (0, 1)$  for the observed time series of microposts is a sufficient feature that Twitter is in the SupC state. Such a state appears from time  $t_c$  on the network evolution process and continues to exist during the time interval  $\Delta t_{\text{SupC}}$ . Starting from time  $t_c$ , the network's behavior becomes unpredictable: avalanches of microposts of any size can appear over time  $t \geq t_c$ .

It is important that the identification of the SupC state of the network does not require a detailed analysis of the interactions between its users at the microlevel. Only an analysis of the time series of microposts for being in the spectrum is sufficient, which does not require significant resource costs. Moreover, estimates of  $\beta$  and  $\gamma$  can be obtained independently, for example, using the DFA method.

Prior to the transition of the network in the SOC state at time  $t_c$  and then in the SupC state, it is in the SubC state during the time interval  $\Delta t_{\text{SubC}}$ . If the network is in this state, then PDF and ACF for the observed time series of microposts are not scale-invariant (see dependencies (1) and (4), respectively). This state can be characterized by exponential laws for PDF and ACF, and the exponent  $\beta$  of PSD has a



value close to 0. It is important to note, it is possible that a network in this state can arise a critical state over time. Before transition in the SOC state, there will be a slow increase in the size of avalanches of microposts over time until a large splash appears in the time series at  $t = +t_c$ . But the appearance of the SOC state on the network is not necessary at all: the network can continue to be in the SubC state until it is completely relaxed due to the loss of user interest in the discussed topic.

An approach to monitor the social network state based on the spectrum analysis, for example, can be effective for identifying the origin of protest movements for which Twitter is one of the tools. In addition, the approach can be used to study the activity of users on the network related to political elections. For example, if the social network is in the SubC state and for the corresponding time series of microposts it is possible to find the interval  $\Delta t \in \Delta t_{\text{SubC}}$ , in which a slow increase in the size of avalanches is observed, then the existence of such  $\Delta t$  is a possible precursor of the appearance of the SOC state and further transition to the SupC state. Another situation is possible. If it is possible to find a relatively small interval  $(t_c, t) \in \Delta t_{\text{SubC}}$ , then the existence of such an interval is a possible precursor of the unpredictability of the behavior on the social network. Starting from time  $t$ , avalanches of microposts of all sizes will appear. Note that all this is nothing more than a discussion of possible applications. To conduct such studies, it is necessary to develop and test algorithms for detecting such integrals, but this is beyond the scope of our study.

The proposed phenomenological social network self-organization equation (9) in the adiabatic approximation (14) describes not only the functioning of Twitter in the SubC, the SOC, and the SupC states, but also the conditions for the transition from one critical state to another. The latter is clearly demonstrated in Table 2.

SubC state of Twitter is a typical social network state. Indeed, the network consists of a large number of users  $\mathcal{N}$ ; each of which follows its own strategy, not related to the strategies of other network users. The intensity of fluctuations of each of the degrees of freedom is commensurate ( $I_\eta = I_h = I_S$ ), and the indicator of the intensity of feedback ( $\epsilon$ ) is equal to 1. The network that is pumping by strategically oriented users ( $S_0 > 0$ ) does not change the network functioning mode qualitatively.

A necessary condition for the network self-organization in a critical state is the appearance of a certain number of users  $S$  within it, who follow a certain strategy, i.e., acting in concert, involving random users as their subscribers. This is an instantaneous unstable state of the network that does not require pumping ( $S_0 = 0$ ). In this state, the intensity of stochastic interactions between SOUs increases significantly ( $I_S \gg I_\eta, I_h$ ), and the streamlining process affects Twitter's self-consistent behavior more strongly ( $\epsilon \in (0, 1)$ ) than if Twitter was in the SubC state.

As a result of pumping of the network by strategically oriented users ( $S_0 > 0$ ), the network moves to the SupC state. In this state, the intensity of stochastic interactions between SOUs is  $I_S \gg I_\eta, I_h$  and  $\epsilon \in (0, 1)$ .

TABLE 2: Critical states of Twitter.

State	$I_i$	$\epsilon$	$S_0$
SubC	$I_\eta = I_h = I_S$	1	$>0$
SOC	$I_S \gg I_\eta, I_h$	(0, 1)	0
SupC	$I_S \gg I_\eta, I_h$	(0, 1)	$>0$

It is fundamentally important that self-organization in a critical state occurs as a result of the agreed action of a relatively small number of users following a single strategy. Random users can not form avalanches of microposts of all sizes on the network.

## 6. Conclusion

In conclusion, we formulate important questions, the answers to which cannot be gotten in the analysis of the phenomenological model (12) of Twitter self-organization in the adiabatic approximation, and we also indicate the possible ways to find the solution.

When discussing Twitter's self-organization mechanisms in a critical state (see Section 2.1), it was noted that one of the possible mechanisms of emergence of the SOC state is the formation of a hierarchical structure in the network, through which avalanches of microposts of all sizes appear in the network. The following is a brief overview of research on this.

Moriano and co-authors state in their paper [64] that "Global events trigger viral information cascades that easily cross community boundaries and can thus be detected by monitoring intra- and inter-community communications." They showed, when a global event (Boston Marathon bombing) occurs, it spreads virally, crossing community boundaries and producing more intercommunity. Despite the fact that we are talking about information cascades, there is no evidence in the article of the self-organized critical nature of this phenomenon. The need of existence of hierarchical structure (user, his subscribers, subscribers of their subscribers, etc.) for the emergence of self-organized critical states is presented in [65, 66]. Morse and co-authors [67] consider the persistent cascades, i.e., recurring patterns of communication among individuals, and relate them to hierarchical spreading of content, analogously to what we discuss in our study. Liu and co-authors [68] devise an embedding model which exploits multiple relations of hashtag-hashtag, hashtag-tweet, tweet-word, and word-word relations based on the hierarchical heterogeneous network.

Stella and co-authors [69] detect power-law relationships between cascade rate and size on Twitter during a voting event and they show how social bots used by human users were capable of creating avalanches of microposts. They showed, "online social interactions during a massive voting event can be used to build an accurate map of real-world political parties and electoral ranks for Italian elections in 2018." They provided, "evidence that information flow and collective attention are often driven by a special class of highly influential users, who exploit thousands of automated agents, also known as bots, for enhancing their online

influence.” In addition, they showed, “influential users generate deep information cascades in the same extent as news media and other broadcasters, while they uniformly infiltrate across the full range of identified groups.” Obviously, highly influential users, who exploit thousands of automated agents are exactly SOUs. Rizoïu and co-authors state in their paper [70] that “Socialbots is more active on Twitter—starting more retweet cascades and retweeting more—but they are 2.5 times more influential than humans, and more politically engaged.” There are studies in which the role of social bots in the emergence of information cascades is considered (e.g., see the works [71, 72]).

González-Bailón and co-authors [73] studied “recruitment patterns in the Twitter network and find evidence of social influence and complex contagion.” They identified the network position of early participants (i.e., the leaders of the recruitment process) and of the users who acted as seeds of message cascades (i.e., the spreaders of information). They found that early participants cannot be characterized by a typical topological position, but spreaders tend to be more central in the network.

Finally, let us consider who can act as SOUs or ROUs. It should be noted that for the identification of SOUs and/or ROUs, time series analysis of microposts is not sufficient. A more meaningful such as sentimental data analysis is beyond the scope of our study.

In the first class, as we discussed previously, there are SOUs. In the cases of political elections, SOUs could be network political bots or “botnets,” who act in concert and use Twitter as a platform for the formation of the avalanche of microposts. For example, Kollanyi and co-authors [74], who analyzed Twitter dataset on US Presidential Election 2016, found that “political bot activity reached an all-time high for the 2016 campaign.” In the cases of protest movements, SOUs, such as leaders or organizers of protests, can be coordinated users who use Twitter as a platform to encourage others to protest. Wang and Caskey indicated in [75] that “Twitter is a tool primarily used for sharing objective, logistical information, along with opinions, to create a unified community and mobilize individuals to participate in a physical space of protests.” And finally, in the cases of natural disasters, we assume SOUs can be the most active, but noncoordinated Twitter users. For example, these users can use Twitter to spread the information about the current state of the environment in their neighborhood, the remaining water in the nearest groceries, etc.

However, it is more interesting to consider the possible users’ nature in the second class. As our analysis showed, there are only ROUs in these datasets:

- (1) End of Term 2016 US Government Twitter Archive. The original tweets were made by 3000 users who are connected with federal US government agencies. We assume that these users acted not in agreement with each other and posted general content. There was no consistent behavior in their action, such as collusion. Therefore, they were ROUs.
- (2) Winter Olympics 2018. Obviously, a lot of different users used Twitter as the platform to advertise the

content; however, all the users pursued their own interests. For example, each sportsman used Twitter as the advertising platform of his brand. In this case, they were ROUs.

- (3) US Government. The original tweets were made by 3400 users who are connected with federal US government agencies. As it was previously, these tweets had general content and were not unified by a common goal, so these users were ROUs.
- (4) News Outlet. The basic tweets were made by news agencies. However, each agency has its own subject, as well as its own way of presenting news. Moreover, each agency promotes news in different time and, sometimes, supports different sides of conflict, for example. We suppose this could be the most reasonable description why these news agencies were ROUs.
- (5) 115th US Congress. The original tweets were made by 535 congress members and their official representatives. As it was in Winter Olympics 2018 example, each member used Twitter as the platform to share his ideas, but they were not unified by a common goal. In this case, they were ROUs.

## Data Availability

Previously reported Tweet Ids data were used to support this study and are available at [<https://doi.org/10.7910/DVN/PDI7IN>, <https://doi.org/10.7910/DVN/5ZVMOR>, <https://doi.org/10.7910/DVN/TQBLWZ>, <https://doi.org/10.7910/DVN/QRKIBW>, <https://doi.org/10.7910/DVN/5CFLJ>, <https://doi.org/10.7910/DVN/DVLJTO>, <https://doi.org/10.7910/DVN/YMJPFJ>, <https://doi.org/10.7910/DVN/2N3HHD>, <https://doi.org/10.7910/DVN/2FIFLH>, <https://doi.org/10.7910/DVN/AEZPLU>, <https://doi.org/10.7910/DVN/UIVHQR>, and <https://doi.org/10.7910/DVN/PYCLPE>]. These prior studies (and datasets) are cited at relevant places within the text as references [43–54].

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was partially funded by the Russian Foundation for Basic Research (grant 16-07-01027).

## References

- [1] P. Sen and B. Chakrabarti, *Sociophysics: An Introduction*, Oxford University Press, Oxford, UK, 2013.
- [2] R. Kutner, M. Ausloos, D. Grech, T. Di Matteo, C. Schinckus, and H. E. Stanley, “Econophysics and sociophysics: their milestones & challenges,” *Physica A: Statistical Mechanics and its Applications*, vol. 516, pp. 240–253, 2019.
- [3] M. Newman, “The physics of networks,” *Physics Today*, vol. 61, no. 11, pp. 33–38, 2008.

- [4] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Critical phenomena in complex networks," *Reviews of Modern Physics*, vol. 80, no. 4, pp. 1275–1335, 2008.
- [5] P. Fronczak, A. Fronczak, and J. A. Holyst, "Phase transitions in social networks," *The European Physical Journal B*, vol. 59, no. 1, pp. 133–139, 2007.
- [6] B. C. Eu, "Thermodynamics of irreversible processes non-equilibrium statistical mechanics, fundamental theories of physics (an international book series on the fundamental theories of physics: their clarification)," *Development and Application*, vol. 93, pp. 12–54, 1998.
- [7] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality: an explanation of the 1/f noise," *Physical Review Letters*, vol. 59, no. 4, pp. 381–384, 1987.
- [8] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Physical Review A*, vol. 38, no. 1, pp. 364–374, 1988.
- [9] D. Markovic and C. Gros, "Power laws and self-organized criticality in theory and nature," *Physics Reports*, vol. 536, pp. 41–74, 2015.
- [10] B. Tadic, M. M. Dankulov, and R. Melnik, "Mechanisms of self-organized criticality in social processes of knowledge creation," *Physical Review E*, vol. 96, Article ID 032307, 2017.
- [11] Q. K. Meng, "Self-organized criticality in small-world networks based on the social balance dynamics," *Chinese Physics Letters*, vol. 28, Article ID 118901, 2011.
- [12] P. A. Noel, C. D. Brummitt, and R. M. D'Souza, "Bottom-up model of self-organized criticality on networks," *Physical Review E*, vol. 89, Article ID 012807, 2014.
- [13] A. Mollgaard and J. Mathiesen, "Emergent user behavior on twitter modelled by a stochastic differential equation," *PLoS One*, vol. 10, Article ID e0123876, 2015.
- [14] M. Aguilera, I. Morer, X. Barandiaran et al., "Quantifying political self-organization in social media. Fractal patterns in the Spanish 15M movement on twitter," in *Proceedings of the 12th European Conference on Artificial Life*, pp. 395–402, Detroit, MI, USA, September 2013.
- [15] L. Kirichenko, V. Bulakh, and T. Radivilova, "Fractal time series analysis of social network activities," in *Proceedings of the IEEE 4th International Scientific-Practical Conference Problems of Infocommunications*, pp. 456–459, Kharkiv, Ukraine, 2017.
- [16] T. De Bie, J. Lijffijt, C. Mesnage C et al., "Detecting trends in twitter time series," in *Proceedings of the IEEE 26th International Workshop on Machine Learning for Signal Processing*, pp. 1–6, Salerno, Italy, 2016.
- [17] D. R. Bild, Y. Liu, R. P. Dick et al., "Aggregate characterization of user behavior in Twitter and analysis of the retweet graph," *ACM Transactions on Internet Technology*, vol. 15, no. 4, 2015.
- [18] C. Remy, N. Pervin, F. Toriumi et al., "Information diffusion on twitter: everyone has its chance, but all chances are not equal," in *Proceedings of the IEEE International Conference on Signal-Image Technology and Internet-Based Systems*, Kyoto, Japan, 2013.
- [19] J. P. Gleeson and R. Durrett, "Temporal profiles of avalanches on networks," *Nature Communications*, vol. 8, p. 1227, 2017.
- [20] C. Liu, X.-X. Zhan, Z.-K. Zhang, G.-Q. Sun, and P. M. Hui, "How events determine spreading patterns: information transmission via internal and external influences on social networks," *New Journal of Physics*, vol. 17, no. 11, Article ID 113045, 2015.
- [21] P. Bak, *How Nature Works. The Science of Self-Organized Criticality*, Springer-Verlag, New York, NY, USA, 1996.
- [22] R. Alvarez, D. Garcia, Y. Moreno et al., "Sentiment cascades in the 15M movement," *EPJ Data Science*, vol. 4, no. 6, 2015.
- [23] S. Pramanik, Q. Wang, M. Danisch et al., "Modeling cascade formation in twitter amidst mentions and retweets," *Social Network Analysis and Mining*, vol. 7, pp. 1–41, 2017.
- [24] M. M. Uddin, M. Imran, and H. Sajjad, "Understanding types of users on twitter," in *Proceedings of the 6th ASE International Conference in Social Computing*, Stanford, CA, USA, 2014.
- [25] A. I. Olemskoi, "Hierarchical pattern of superdiffusion," *Journal of Experimental and Theoretical Physics Letters*, vol. 71, no. 7, pp. 285–288, 2000.
- [26] E. Bakshy, A. Arbor, W. A. Mason et al., "Everyone's an influencer: quantifying influence on twitter," in *Proceedings of the WSDM'11 Fourth ACM International Conference on Web Search and Data Mining*, pp. 65–74, Hong Kong, China, 2011.
- [27] E. Himmelboim, M. A. Smith, L. Rainie et al., *Classifying Twitter Topic-Networks Using Social Network Analysis*, Social Media + Society, New York, NY, USA, 2017.
- [28] D. Antoniadis and C. Dovrolis, "Co-evolutionary dynamics in social networks: a case study of Twitter," *Computational Social Networks*, vol. 2, pp. 1–21, 2015.
- [29] J. Portugali, "Self-organization and the city," in *Encyclopedia of Complexity and Systems Science*, R. Meyers, Ed., Springer, New York, NY, USA, 2009.
- [30] J. W. Kantelhardt, "Fractal and multifractal time series," in *Mathematics of Complexity and Dynamical Systems*, R. Meyers, Ed., Springer, New York, NY, USA, 2012.
- [31] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.
- [32] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [33] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, "Cut-offs and finite size effects in scale-free networks," *European Physical Journal B*, vol. 38, pp. 205–209, 2004.
- [34] J. S. Andrade, H. J. Herrmann, R. F. Andrade et al., "Apollonian networks: simultaneously scale-free, small world, euclidean, space filling, and with matching graphs," *Physical Review Letters*, vol. 94, pp. 1870–1871, 2005.
- [35] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the internet: a source of enormous confusion and great potential," *Notices of the AMS*, vol. 56, pp. 586–599, 2009.
- [36] A. Mislove, M. Marcon, K. P. Gummadi et al., "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, San Diego, CA, USA, 2007.
- [37] H. Thomas, J. M. Read, L. Danon et al., "Testing the hypothesis of preferential attachment in social network formation," *EPJ Data Science*, vol. 4, p. 13, 2015.
- [38] A. Barabasi, *Network Science*, Cambridge University Press, Cambridge, UK, 2016.
- [39] M. P. H. Stumpf and M. A. Porter, "Critical truths about power laws," *Science*, vol. 335, no. 6069, pp. 665–666, 2012.
- [40] M. O. Jackson and B. W. Rogers, "Meeting strangers and friends of friends: how random are social networks?," *American Economic Review*, vol. 97, no. 3, pp. 890–915, 2007.
- [41] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature Communications*, vol. 10, p. 1017, 2019.
- [42] M. Gerlach and E. G. Altmann, "Testing statistical laws in complex systems," *Physical Review Letters*, vol. 122, Article ID 168301, 2019.
- [43] J. Littman, L. Wrubel, and D. Kerchner, "United States presidential election tweet Ids," 2016.
- [44] J. Littman and S. Park, "Women's march tweet Ids," 2017.



- [45] J. Littman, D. Kerchner, and L. Wrubel, "End of term 2016 U.S. government twitter archive," 2017.
- [46] J. Littman, "Hurricanes harvey and irma tweet Ids," 2017.
- [47] J. Littman, "Immigration and travel ban tweet Ids," 2018.
- [48] J. Littman, "Charlottesville tweet Ids," 2018.
- [49] J. Littman, "Winter olympics 2018 tweet Ids," 2018.
- [50] J. Littman, D. Kerchner, and L. Wrubel, "U.S. government tweet Ids," 2017.
- [51] J. Littman, L. Wrubel, D. Kerchner et al., "News outlet tweet Ids," 2017.
- [52] L. Wrubel, J. Littman, and D. Kerchner D, "U.S. congressional election tweet Ids," 2019.
- [53] J. Littman, "115th U.S. congress tweet Ids," 2017.
- [54] J. Littman, "Ireland 8th tweet Ids," 2018.
- [55] H. E. Hurst, "Long term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–799, 1951.
- [56] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E*, vol. 49, no. 2, pp. 1685–1689, 1994.
- [57] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 2010.
- [58] S. V. Buldyrev, A. L. Goldberger, S. Havlin et al., "Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis," *Physical Review E*, vol. 51, no. 5, pp. 5084–5091, 1995.
- [59] A. I. Olemskoi and A. V. Khomenko, "Three-parameter kinetics of a phase transition," *Journal of Theoretical and Experimental Physics*, vol. 83, pp. 1180–1192, 1996.
- [60] A. I. Olemskoi, "Theory of stochastic systems with singular multiplicative noise," *Physica-Uspekhi*, vol. 41, no. 3, pp. 269–301, 1998.
- [61] A. I. Olemskoi, A. V. Khomenko, and D. O. Kharchenko, "Self-organized criticality within fractional Lorenz scheme," *Physica A: Statistical Mechanics and its Applications*, vol. 323, pp. 263–293, 2003.
- [62] D. Ruelle and F. Takens, "On the nature of turbulence," *Communications in Mathematical Physics*, vol. 20, no. 3, pp. 167–192, 1971.
- [63] A. Dmitriev, V. Kornilov, and S. Maltseva, "Complexity of a microblogging social network in the framework of modern nonlinear science," *Complexity*, vol. 2018, Article ID 4732491, 2018.
- [64] P. Moriano, J. Finke, and Y. Ahn, "Community-based event detection in temporal networks," *Scientific Reports*, vol. 9, p. 4358, 2019.
- [65] J. P. Gleeson, J. A. Ward, K. P. O'Sullivan et al., "Competition-induced criticality in a model of meme popularity," *Physical Review Letters*, vol. 112, Article ID 048701, 2014.
- [66] J. P. Gleeson, K. P. O'Sullivan, R. A. Baños, and Y. Moreno, "Effects of network structure, competition and memory time on social spreading phenomena," *Physical Review X*, vol. 6, Article ID 021019, 2016.
- [67] S. Morse, M. C. González, and N. Markuzon, "Role of persistent cascades in diffusion," *Physical Review E*, vol. 99, pp. 12323–12331, 2019.
- [68] J. Liu, Z. He, and Y. Huang, "Hashtag2Vec: learning hashtag representation with relational hierarchical embedding model," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 3456–3462, Stockholm, Sweden, July 2018.
- [69] M. Stella, M. Cristoforetti, and M. De Domenico, "Influence of augmented humans in online interactions during voting events," *PLoS One*, vol. 14, Article ID e0214210, 2019.
- [70] M.-A. Rizoio, T. Graham, R. Zhang et al., "DEBATENIGHT: the role and influence of socialbots on twitter during the first 2016 U.S. Presidential debate," in *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, pp. 300–309, Palo Alto, CA, USA, June 2018.
- [71] M. T. Bastos and D. Mercea, "The brexit botnet and user-generated hyperpartisan news," *Social Science Computer Review*, vol. 27, 2017.
- [72] C. Shao, G. L. Ciampaglia, O. Varol et al., "The spread of low-credibility content by social bots," *Nature Communications*, vol. 9, p. 4787, 2018.
- [73] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, "The dynamics of protest recruitment through an online network," *Scientific Reports*, vol. 1, pp. 1–7, 2011.
- [74] B. Kollanyi, P. N. Howard, and S. C. Woolley, *Bots and Automation over Twitter during the U.S. Election, Data Memo*, Project on Computational Propaganda, Oxford, UK, 2016.
- [75] Z. Wang and K. Caskey, "#Occupywallstreet: an analysis of twitter usage during a protest movement," *Social Networking*, vol. 5, no. 4, pp. 101–117, 2016.



