

**Mechanisms of self-organized criticality in social processes of knowledge creation**Bosiljka Tadić,<sup>1</sup> Marija Mitrović Dankulov,<sup>2</sup> and Roderick Melnik<sup>3</sup><sup>1</sup>*Department of Theoretical Physics, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*<sup>2</sup>*Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia*<sup>3</sup>*MS2Discovery Interdisciplinary Research Institute, M2NeT Laboratory and Department of Mathematics, Wilfrid Laurier University, Waterloo, Ontario, Canada, N2L 3C5*

(Received 17 May 2017; revised manuscript received 4 August 2017; published 5 September 2017)

In online social dynamics, a robust scale invariance appears as a key feature of collaborative efforts that lead to new social value. The underlying empirical data thus offers a unique opportunity to study the origin of self-organized criticality (SOC) in social systems. In contrast to physical systems in the laboratory, various human attributes of the actors play an essential role in the process along with the contents (cognitive, emotional) of the communicated artifacts. As a prototypical example, we consider the social endeavor of knowledge creation via Questions and Answers (Q&A). Using a large empirical data set from one of such Q&A sites and theoretical modeling, we reveal fundamental characteristics of SOC by investigating the temporal correlations at *all scales* and the role of cognitive contents to the avalanches of the knowledge-creation process. Our analysis shows that the universal social dynamics with power-law inhomogeneities of the actions and delay times provides the primary mechanism for self-tuning towards the critical state; it leads to the long-range correlations and the event clustering in response to the external driving by the arrival of new users. In addition, the involved cognitive contents (systematically annotated in the data and observed in the model) exert important constraints that identify unique classes of the knowledge-creation avalanches. Specifically, besides determining a fine structure of the developing knowledge networks, they affect the values of scaling exponents and the geometry of large avalanches and shape the multifractal spectrum. Furthermore, we find that the level of the activity of the communities that share the knowledge correlates with the fluctuations of the innovation rate, implying that the increase of innovation may serve as the active principle of self-organization. To identify relevant parameters and unravel the role of the network evolution underlying the process in the social system under consideration, we compare the social avalanches to the avalanche sequences occurring in the field-driven physical model of disordered solids, where the factors contributing to the collective dynamics are better understood.

DOI: [10.1103/PhysRevE.96.032307](https://doi.org/10.1103/PhysRevE.96.032307)**I. INTRODUCTION**

In recent years, the self-organized criticality (SOC) is considered as one of the principal mechanisms responsible for the emergence of new features at a larger scale in various complex systems. The transition from the microscopic interactions to the collective behavior involves nonlinear dynamical phenomena when the system is driven out of equilibrium (for an overview of physical systems exhibiting SOC, see recent reviews in [1–3] and the references there). In this context, SOC refers to the dynamical self-organization among the interacting units in response to repeatedly applied infinitesimal driving; the system's adaptation to the driving force leads to robust metastable states with system-wide correlations, fractal dynamics, and avalanches as the key signatures of criticality [2–6]. In this context, an avalanche is recognized as a mesoscopic dynamical structure consisting of a sequence of connected elementary events (a precise definition is given in Sec. II). It has been newly pointed out that SOC plays a role in the functioning of biological [7] and diverse other complex systems from neuronal dynamics [8,9] to animal behavior [10] and human history [11]. For instance, the analysis of vast amounts of the available brain imaging data and theoretical modelings provided the evidence that supports SOC as an underlying mechanism of the brain functional stability [8,12].

Although the avalanching behavior and other signatures of criticality are readily observable in the empirical data of online social dynamics [13–19], much less attention has

been devoted to understanding the origin and the precise role of SOC in social systems. The key open question is whether the social avalanches represent a unique class of self-organized phenomena or, otherwise, they can be reduced to standard models of physical SOC systems, describing the transition from the microscopic interactions to the observed complex spatiotemporal patterns. Another interesting aspect of the problem concerns the interplay of the coevolving network structure and the social dynamics that it supports. The question whether the SOC process shapes the structure, or the network evolution enables the self-tuning towards the criticality remains open. Here, we address these issues by analysis of the empirical data of knowledge-creation social endeavors and using theoretical modeling.

In physics, striking examples of the multiscale dynamics characterizing SOC are observed in the turbulent flow [2,20–22] and the kinetics of earthquakes [23,24]. The signatures of SOC are also found in experiments with stressed granular materials [25,26], driven disordered systems at a hysteresis loop [27–30], and porous shape-memory alloys [31,32]. Furthermore, the avalanching dynamics is characteristic of the conduction in the assembled networks of living neurons in a solvent [33] and nanoparticle films with single-electron tunnelings conduction [34], as well as to the motion of topological objects, such as vortices [35,36] and domain walls [37–39]. The theoretical concepts were developed to describe the emergence of collective behaviors

from the microscopic interactions among many constitutive elements both in the classical and quantum systems [40]. In this regard, often paradigmatic models were used as the instruments for investigations. Further, the use of scaling and renormalization group ideas provided a better understanding of the role of different scales in the dynamical systems driven away from the equilibrium [41–43]. The precise description of the interactions in these physical systems allows investigating the microscopic mechanisms responsible for the triggering and propagation of an avalanche. The standard feature of all SOC systems is the accumulation of free energy, which then dissipates through the avalanches. While the energy source and triggering mechanisms are physics specific, common to all SOC systems is that the avalanching response is not in proportion to the forcing. Consequently, the power-law distributions of the avalanches appear as one of the key features of SOC states. It has been recognized that the propagation of avalanches involves three phases [2]: the initial growth phase is supported by multiplicative chain reactions until the maximum dissipation is reached in the peak period, after which the activity is reduced and eventually diminishes, in the stopping phase. Often, the essence of SOC can be captured by the elementary dynamics of sand-pile automata [4,5] on a two-dimensional lattice. In real systems, however, the presence of many physical parameters that can influence the dynamics makes it difficult to distinguish the potential SOC states (attractor with a large basin of attraction) from the dynamical phase transition, which occurs by fine tuning of a relevant parameter.

In this work, we investigate the nature of avalanches in the prototypical human collaborative endeavor of knowledge creation [44]. In this process, the knowledge and expertise of individual actors are transferred into a social value [45,46]—the collective knowledge, which is shared by all participants in the process. In contrast to the physical systems in the laboratory, the human cooperation, as well as the new collective states, are evident [44,47]. Therefore, the empirical data on these social systems represent a valuable source to investigate the origin of self-organized criticality. On the other hand, certain attributes of the human participants are crucial to the social cooperation; they remain elusive to the accurate theoretical modeling of the interactions, which underlie the avalanche formation. In particular, the process of knowledge creation requires the appropriate expertise of the participating actors among other human attributes. Thus, the subdynamics representing the use of the communicated cognitive contents tends to constrain the social process itself. In this regard, it remains unclear how these different aspects (social and cognitive) of the dynamics contribute to the appearance and propagation of the avalanches. To address this question, we combine the analysis of the empirical data from Questions and Answers (Q&A) site Mathematics Stack Exchange (<https://math.stackexchange.com/>) with the agent-directed modeling; the cognitive contents of each artifact are systematically encoded in the considered data sets. The agent's attributes are statistically similar to the users in the data, while their expertise is varied. The system is driven by the arrival of new agents. Our analysis reveals that the occurrence of avalanches is a robust social phenomenon, whereas their fine structure, geometrical, and fractal characteristics are

affected by the distribution of the expertise over the actors. Furthermore, the interplay between the social and knowledge processes is fueled by the constant tendency towards the expansion of innovation. For a deeper understanding of the potential mechanisms, we use a comparison with the better controlled avalanching dynamics in physical systems. For this purpose, we analyze the model of a disordered system of the interacting spins, where the critical states at the hysteresis loop appear in the interplay between the driving by the applied magnetic field and the domain-wall pinning along the implanted magnetic (soft) and structural (hard) defects. Although an analogy between spin alignments can be extended to knowledge matching among the social subjects, our objective here is different. We compare the fractal features of the avalanche sequences in both systems, which appear to be similar in a particular range of parameters of the physical system. These comparisons permit us to identify certain factors of the social dynamics that are essential to the appearance of the collective state and can motivate further research towards a viable modeling of the social self-organization.

In the following Sec. II, we introduce the essential characteristics of the processes of Q&A and describe details of the agent-based model and the structure of the bipartite network that coevolves with the social interaction. Then Sec. III presents a detailed analysis of the knowledge-bearing avalanches both from the empirical data and simulations. The simulations and analysis of the avalanches in the driven spin system are given in Sec. IV. A summary of the results and the discussion are given in Sec. V.

## II. THE STOCHASTIC PROCESS OF KNOWLEDGE CREATION AND THE COEVOLVING NETWORKS

The knowledge creation via Questions & Answers is a collaborative social endeavor, in which the knowledge of each participant is shared with others. By its nature, the interaction between these participants is indirect, mediated by questions and answers, in a way similar to user interactions via posted texts on blogs [13,48,49]. Thus, the environment of the knowledge sharing can be represented as a coevolving *bipartite network* with the *actors* (users, agents) as one partition, and the *artifacts* (questions, answers) as the other partition [44]. In epistemology, to create a common value (knowledge), *meaningful social interactions* are required, in which the actor's response is adequate to the needs of others [45,46]. Specifically, the actor possessing an expertise can meaningfully act on the artifact where this particular expertise is required. Thus, the essence of the dynamics is the contents-matching rule, as schematically illustrated in Fig. 1(a). A part of the evolving network extracted from the empirical data is also shown in Fig. 1(b). Hence, the knowledge-creation process consists of two mutually interconnected factors: the social dynamics and the constrained use of the cognitive contents. The strict use of the available expertise in the knowledge-creation processes is in marked contrast with the informal social communications on blogs and similar systems, where the user's natural interests and emotions drive the activity [13].

In the data from the analyzed Q&A site, the cognitive content of each artifact is encoded by up to five tags, according

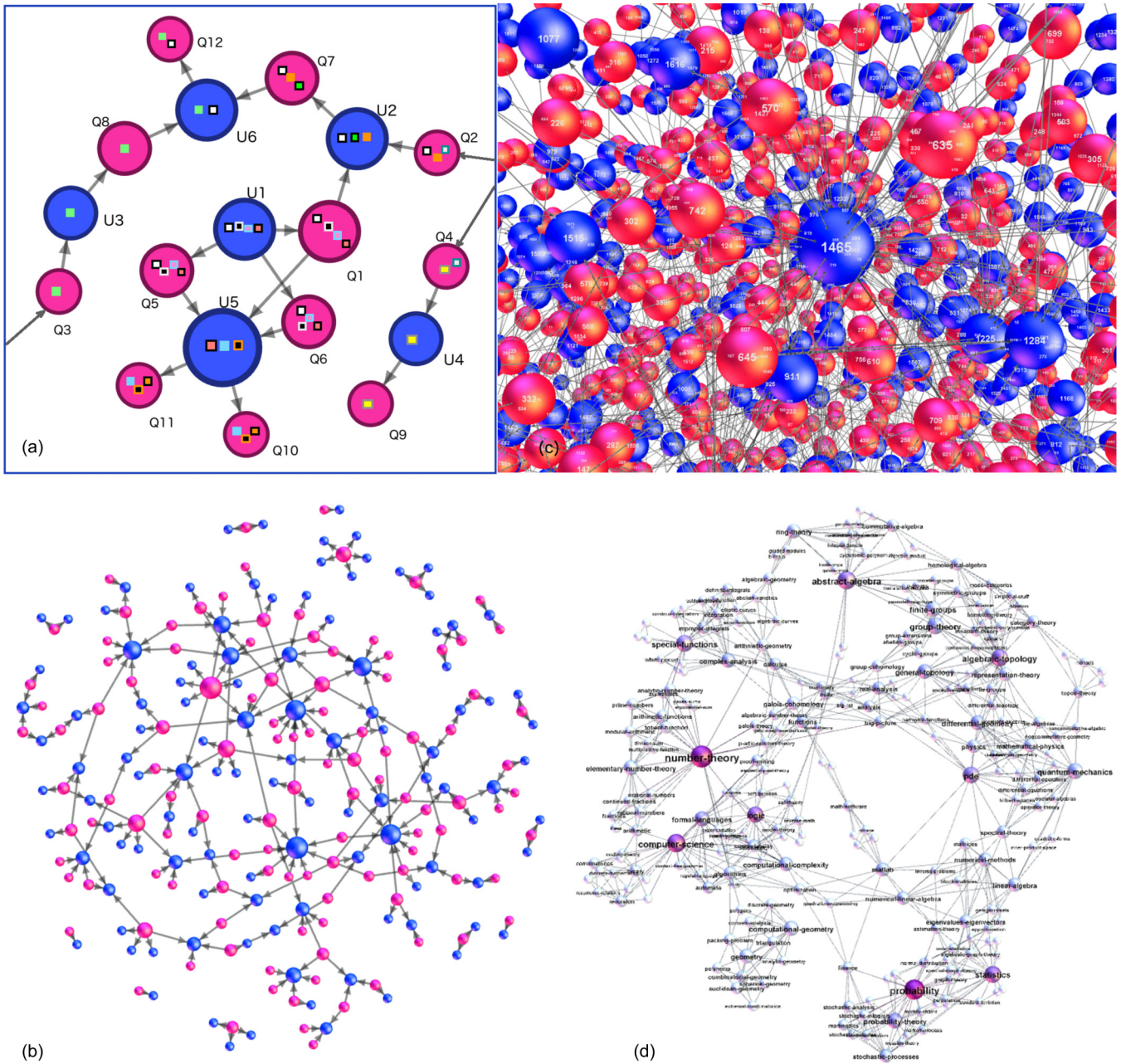


FIG. 1. (a) Schematic view of the knowledge exchange via Q&A on a bipartite network of actors (blue nodes) and questions and answers (red nodes); boxes of different colors represent the cognitive contents (tags). A directed link from question to actor node indicates activity on that question, where the matching (at least in one tag) between the actor’s expertise and the contents of the question is required; the outgoing link represents posting the question or answer, whose content contains the actor’s knowledge. Panels (b),(c) show two bipartite networks extracted from the empirical data; the user nodes (blue) connect to question nodes (red) which compress all existing answers to that question, and the direction of the link indicates the question on which the user was active either by posting or answering it. Specifically, (b) the innovation layer in the evolving bipartite network extracted at the end of year one, consists of the recently active questions and the users whose activity on these questions occurred in the last  $T_0 = 100$  min and the nodes to which they were connected within the time depth of  $6T_0$ . (c) A close up of the compressed bipartite network of users and the questions filtered such to contain the tag “Linear Algebra” among other tags; the network represents the activity within the first two months of the considered empirical data. (d) The explicit-knowledge network containing the innovation tags attached to the tags of year one.

to the standard mathematical classification scheme (MCS), for example, “Graph Theory,” “Probability,” “Stochastic Processes,” “Linear Algebra,” “Algebraic Topology,” “Differential Geometry,” and others. Whereas, the information about the user’s expertise participating in the process can be inferred

statistically as described in methods of [44]; see the inset to Fig. 2(a). This figure suggests a rather broad distribution of the expertise (i.e., different number of tags  $2^{E_i}$ ) over the users in the native system. The main Fig. 2(a) shows another key characteristic of the experimental system: the

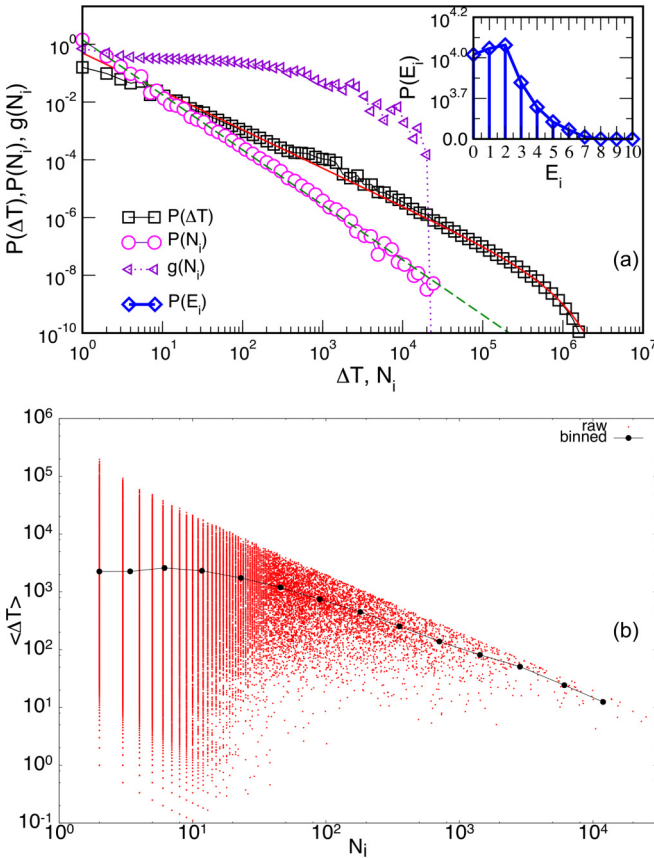


FIG. 2. (a) The distribution  $P(N_i)$  of the number of actions  $N_i$  per user, the dependence  $g(N_i)$ , and the delay time  $P(\Delta T)$  of the users in the empirical data set. Inset: The distribution of entropy  $P(E_i)$  based on the probability of the users' expertise; the data are extracted from the empirical set and used for designing the attributes of the agents. (b) Scatter plot of the user-averaged delay times  $\langle \Delta T \rangle$  against the number of actions  $N_i$ .

user's heterogeneity in the number of actions  $N_i$  and a broad range of their interactivity times  $\Delta T$ . While the scaling exponents indicated by these histograms are system specific, the prominent power-law decay of both quantities manifests the universally observed characteristics of the human behavior online. Further patterns of the user's activity can be determined from the time stamp. Notably, a significant part of the actions of the present users is directed towards the issues posted by new arrivals. Occasionally, an active user looks for an older question with currently searched contents and brings it to the view of others. The ratio of the posted versus answered questions  $g(N_i)$  was found to depend on the number of actions of a user; cf. Fig. 2(a). The very active users are in the minority; they often respond to the questions. On the other side, the majority of the least active users are engaged in posting questions and, by getting the satisfactory response, they disappear for a longer period. Note that a similar dependence of the delay times on the number of actions applies to the user-averaged delays; cf. Fig. 2(b), but not to each particular user. Certain regularities may exist for groups of users of a similar activity level [16], but they are not of interest in the present context. The arrival of new users (referred to the beginning of the data set) captured by the *time series*  $p(t)$

represents a stochastic process, which depends on the user's off-line life. Here, we adopt the interval of 10 min as a suitable time step for the pace of the activity in the system and use the empirical time series  $p(t)$  and its randomized version, as described in the following text, to create the number of agents per time step in the simulations (on average less than one agent per step). Then the reference time depth is  $T_0 = 10$ .

Observing the requirements for the minimal agent-based model (ABM) of the Web users [50], we have introduced a model where the action rules and the attributes of the agents are taken from these empirical distributions, whereas their expertise can be varied [44]. Here, we briefly describe the main features of the model which is used for the simulations in this work. In particular, in each time step:

(i) *Agents are created.* The number  $p(t)$  of new agents is created; their profiles are defined by the number of actions  $N_i \in P(N_i)$ , the ratio of the posted vs answered questions  $g(N_i)$ , and the expertise (according to the selected distribution); the new agents are placed on the active list;

(ii) *The agent's action performed.* Each agent from the active list either posts a new question or selects one from the list of recently considered artifacts to act on it. The artifacts that are connected to the agent's network neighborhood are looked at first; with a fixed probability (0.5) the agent also finds a related question in the whole network, thus bringing it to the currently active context. In each case, the expertise matching rule applies.

(iii) *The active questions and network are updated.* The list of active questions within time depth  $T_0$  is maintained, and the network connections are updated according to the executed actions; the agents linked to the questions on which the activity occurred within the previous three steps are prompted for a new action.

(iv) *New delay times are determined.* An agent gets a new delay time  $\Delta t \in P(\Delta t)$  after every completed action or after being prompted for a new action.

(v) *The status of each agent is updated.* The number of actions of all agents is updated according to their activity, and the agents, whose number of steps reached the predefined  $N_i$  are removed; the delay times of each remaining agent is updated, and each agent whose delay since the previous event expired is placed on the active agents' list.

The expertise of an agent is a set of tags taken randomly from the list of 32 tags. The considered distributions of the expertise are  $Exp1$  and  $Exp3$ , corresponding to single-tag and three-tag expertises, respectively, and a broad range of the expertise  $ExpS$ , according to the empirical distribution shown in the inset in Fig. 2. For a comparison, we also consider a situation ( $\mu$  process) where, instead of the actions described above in step (ii), an agent finds an artifact in the entire system and acts on it with a fixed probability (0.25) while disregarding the expertise matching rule. Note that, in this case, the expertise matching can occur by chance; the agent's expertise is taken from the distribution  $ExpS$ .

Considering the evolution of the system in [44], we have shown that the process is characterized by the innovation growth with the number of events. In this context, the innovation is suitably defined as the number of new combinations of tags. The innovation is introduced into the system by new arrivals and the actions of the other agents through adding their

expertise to the accumulating contents on different artifacts. The innovation growth was observed both in the empirical data as well as in the simulations [44]. It applies to various distributions of the expertise, excluding the case where all actors possess a strictly single-tag expertise. In this limiting case, the tag-matching rule prevents interlinking with other contents, thus leading to isolated communities that share the same tag. In contrast, the innovation grows whenever at least a single actor possesses a combination of different contents in its expertise. The illustration in Fig. 1(a), for instance, indicates how the green tag of the artifact Q3 remains detached until the actor's U6 expertise combines it with the contents of Q7 and the actor U2 who posted it. The observed pace of the innovation growth depends on the distribution of the expertise (by the fixed activity patterns of the actors) [44]. It is important to stress that, in the empirical data, the combination of tags in the expertise of each user already possesses a logical structure of mathematical knowledge. Hence, through these meaningful interactions that structure is preserved during the process. Consequently, the developing network of the used contents also exhibits the logical structure, as shown in [47] by the community detection in the corresponding networks of tags; an example of such an explicit knowledge network extracted from the same data is shown in Fig. 1(d).

#### A. Growth of the bipartite networks by adding the innovation layers

The interplay between the network structure and the stochastic processes taking part on it consists of the central problem in understanding the networks evolution and their applications in various fields [51–53]. Typically, the graph architecture represents geometrical constraints that shape the diffusionlike processes, likely to cause an anomalous diffusion. For example, the superdiffusion of the information packets [54] occurs on the correlated scale-free network when the traffic rules appropriately utilize the underlying structure. Some unusual situations arise when the structure evolves at the same pace as the SOC avalanching process on it. The random rewiring during the steps of the SOC dynamics has been shown to reduce the avalanche cutoffs [55], thus preventing a catastrophic event to occur. When the rewirings are strictly confined to the current avalanche area, the network appears to have the scale-free degree distribution, where the scaling exponent coincides with one of the avalanche size distributions [56]. Other growth models that apply thresholdlike constraints inspired by SOC dynamics may yield the nonextensive features and scale freeness [57]. In bipartite networks, however, each partition plays a different role in the process, which leads to a more complicated structure–dynamics interplay. As mentioned above, these types of networks often appear in the social dynamics on websites which maintain indirect communication between the users. The appropriate analysis of the artifacts mediating the users revealed [13,44,48,49] how their emotional or cognitive contents affect the network from the node's degree to the mesoscopic community structure.

In the Q&A data set that we consider here, the network growth as well as the pattern of activity of each user and the targeted questions can be extracted from the time stamp in the data. Moreover, to visualize the bipartite networks, we

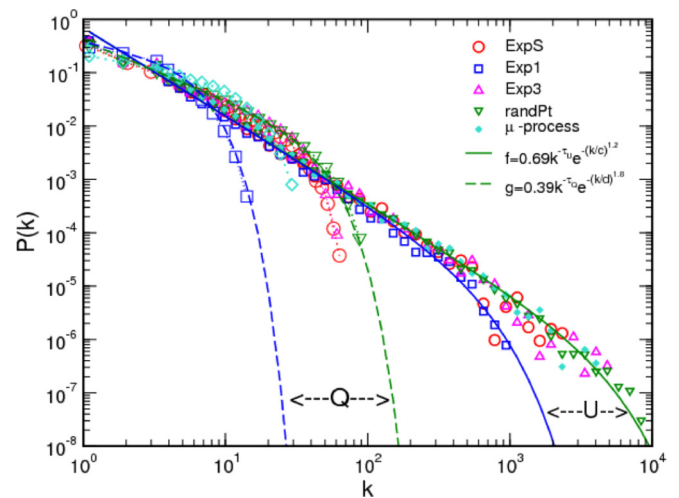


FIG. 3. The degree distribution of the actors and question nodes in the bipartite networks for different expertise and driving indicated in the legend. The power law fits with the exponent  $\tau_U = 1.58 \pm 0.07$  and the stretching cutoffs in the range  $c \in [508, 3920]$  between two solid lines marked by  $\leftarrow U \rightarrow$  are of the actor's degree. The distributions for the corresponding question nodes (indicated by the same type of symbol) have smaller cutoffs  $d \in [5.8, 42]$  between two dashed lines marked by  $\leftarrow Q \rightarrow$  and the exponent  $\tau_Q = 1.09 \pm 0.1$ . In the case of *Exp1* the exponents  $\tau_Q \sim 0.5 \pm 0.16$  apply in a very narrow range, while the exponential distribution fits the data for the  $\mu$  process.

introduce a compressed node which includes the question and all answers related to that question. A particular example of such compressed bipartite network from the empirical data is shown in Fig. 1(c). The corresponding networks from the simulated data exhibit the mesoscopic structure. The structure of communities sharing the emerging knowledge crucially depends on how the expertise is distributed over the involved participants, as it was shown in [44]. Here, we are interested in the statistical properties of nodes in these bipartite networks. The results of the degree distributions are shown in Fig. 3. They can be fitted by the power law with stretched-exponential cutoffs, which allows a comparison of the user properties in Fig. 2(a). Statistically, the degree distribution of the agent's nodes in these networks follows the slope of the predefined number of actions  $P(N_i)$ , as expected, while their cutoffs depend on the expertise of the agents. It is also interesting to point out the constraints due to the power-law decay of the delay time distribution; according to some recent studies [58], it can contribute to the convergence towards scale invariance in the growing systems. In the case of the question partition, the cases where the expertise-matching dominates exhibit a similar law but in a reduced range. In the meantime, the power-law behavior is reduced and the cutoffs dominate in the case *Exp1* and  $\mu$  process (see Fig. 3).

For the purpose of this work, it is interesting to recognize the *innovation growth layer*, see Fig. 1(b), as the segment of the growing bipartite network where the most recent activity occurs. As stated above, the new arrivals potentially bring the new combinations of the knowledge contents in their expertise, which is expressed in the questions and answers. The set of currently active artifacts are posted or answered within a

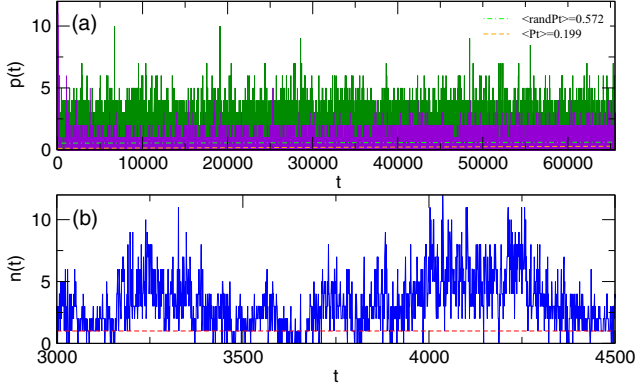


FIG. 4. (a) The driving signals that are used in simulations:  $p(t)$ , front curve, and its randomized version  $randPt$ , back curve. (b) Part of the activity time series from the empirical data, exhibiting some large and small avalanches above the baseline (dashed line).

relatively small time window ( $T_0$ ). These nodes often occur at the outer layer of the network; see Fig. 1(b). Then the currently active users connect to these new artifacts while obeying the expertise matching rule; thus, they connect the new contents to the issues of their previous activity and further through the network of the connected users and their artifacts in a given time depth. Therefore, the recently added artifacts connect the in-depth network via the active users and their previously established connections. Considering a particular time depth, for instance,  $6T_0$ , with the window  $T_0 = 100$  min, we focus on the currently active layer of the network. This layered growth of the bipartite network is fundamentally conditioned by the nature of online human communications, where the latest posts appear on the top. Besides, at each event, an artifact older than the considered time depth is searched with a small probability and connected to the currently active matching contents. The updated active layer then serves as an environment where the next arrivals often attach to, and so on. In the context of open dynamical systems, the addition of new users (or agents) and their artifacts can be seen as the *driving mode* of this bipartite networked system.

### B. Extracting the time series and avalanches from the sequence of events

Our focus in this work is on the avalanching behavior which occurs as the network's response to the driving. Therefore, from the sequence of events in the empirical data or the simulated events, we first construct the corresponding time series that capture the fluctuations of the system's activity over time. The possible occurrence of the clustering of events along these time series is a signature of the avalanching dynamics. For the illustration, an example of the time series with the avalanches is shown in Fig. 4(b). Specifically, an avalanche is identified as a segment of the time series consisting of the data points  $n(t)$  between two consecutive drops of the signal to the baseline, which is set above a zero or the noise level. To define the baseline, we use a standard approach, as described in [27] for the experimental Barkhausen noise signal. When the signal contains an (extrinsic) noise, the baseline is first put as a horizontal line with a maximum number of intersects

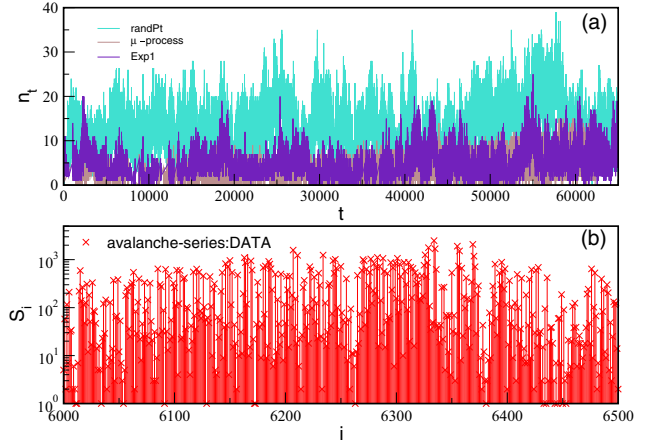


FIG. 5. (a) Examples of time series of the activity simulated by the agent-based model for the different expertise of the agents and  $\mu$  processes. (b) Sequence of the avalanches determined in the empirical data set—a closeup in the segments with large activity.

with the signal, then the part below the line is considered as noise. The standard deviation  $\sigma$  of the noise is then computed, and the baseline is shifted upward by the distance  $\sigma$ . In the simulated quasistatic driving, Sec. IV, the signal drops to the zero level before next driving event occurs, making the zero level a natural baseline (see also the discussion below). The two intersections of the signal with the baseline are recognized as the beginning  $t_b$  and the end  $t_e$  time of the avalanche. Then the avalanche size  $s$  is given by the sum of all data points between the marked beginning and the end of the avalanche while the distance between these two points along the time axis defines the avalanche duration  $T$ , i.e.,

$$s = \sum_{t=t_b}^{t_e} n(t); \quad T = t_e - t_b. \quad (1)$$

The representative examples of the time series studied in this work are shown in Figs. 4 and 5(a). As the example in Fig. 4 shows, the avalanches in the considered stochastic process of knowledge creation differ in size, duration, and shape, closely reflecting the way that the activity propagates in the network. Moreover, a massive avalanche may follow immediately after a small one and vice versa. See also a closeup of the avalanche sequence derived from the empirical data in Fig. 5(b).

It is important to notice that the time series studied in Sec. III are fractal; hence, the avalanches defined through Eq. (1) possess self-affinity. This implies that, by changing a linear scale  $\ell$ , both the segment of the baseline along the  $x$  axis, i.e., the avalanche duration  $T$ , and the corresponding area above the baseline in the  $y$  direction, representing the avalanche size  $s$ , scale with *different exponents* such that the respective dimensionless quantities  $s/\ell^D$  and  $T/\ell^z$  remain unchanged. Consequently, the distributions of the avalanche sizes and durations then obey the following scale invariance [1,2]

$$P(s, \ell) = s^{\tau_s} \mathcal{S}(s/\ell^D); \quad P(T, \ell) = T^{\tau_T} \mathcal{T}(T/\ell^z), \quad (2)$$

where  $\tau_s$  and  $\tau_T$  are the scaling exponents of the avalanche size and duration, respectively, and  $D$  and  $z$  are the corresponding

fractal exponents. Furthermore, for such self-affine objects, the following relation holds between the size of avalanches which have a given duration  $T$  and the duration  $T$

$$\langle s \rangle_T \equiv \sum_{t=t_b}^{t_b+T} n(t) \sim T^{\gamma_{ST}}, \quad (3)$$

with the scaling relation  $\gamma_{ST} = (\tau_T - 1)/(\tau_s - 1)$ . This property of the analyzed avalanches, which we prove in Sec. III, further implies that the scaling exponents are not sensitive to the exact position of the baseline, which discriminates the noise level. Note that we use the initial part of the empirical time series comprising of 65 536 data points, where these methods can be safely applied. In the next section, we quantitatively study these features of the avalanches by considering the time series from both the empirical data and the simulated events.

Figure 5(a) panel shows several examples of the time series of the number of event  $n(t)$ . Specifically, the time series indicated by *Exp1* corresponds to the case when each agent possesses a single-tag expertise and these agents are added with the pace  $p(t)$  as the new users appear in the real system. The time series marked by *randPt* is the system of agents with the distribution of the expertise *ExpS* taken from the empirical data and driven by the randomly shuffled signal  $p(t)$  signal. Apart from a few high values at the start, the signal  $p(t)$  exhibits an increasing trend, which induces larger activity at later times. By randomizing the time series, however, these larger values may occur randomly along the time axis; consequently, the initial part of the signal which is used for simulations appears to be higher than the original  $p(t)$ ; cf. Fig. 4(a). Hence, the underlying network grows faster when the system is driven by *randPt*; on average, the number of added agents putting their artifacts per 1000 steps is 572, compared to 199 in the case of the original signal  $p(t)$ . In this way, this accelerated network growth mimics a larger driving rate in the context of SOC systems. The simulated data consist of 65 536 steps, corresponding to the first 15 months of the real system time, where we find that 13 045 users were active, posting 21 998 questions and 179 537 answers. The bottom panel of the same figure shows the avalanche sequence determined from the empirical data set.

### III. THE STRUCTURE OF AVALANCHES IN KNOWLEDGE-CREATION PROCESSES

The use of fractal geometry and nonlinear analysis of time series has advanced the understanding of complex systems. Here, we employ the detrended multifractal analysis to study the time series of events as well as the sequences of the avalanches (clustered events) for various model parameters and the empirical data of the knowledge-creation processes. By investigating these complex time series at all scales, we aim to reveal a fine structure of the underlying SOC states of the system.

#### A. Temporal correlations and avalanche sizes

The occurrence of avalanches in composite signals is not accidental but built on the temporal correlations at a larger scale. These correlations are manifested in the corresponding power spectrum as a power-law decay  $W(\nu) \sim \nu^{-\phi}$ , for a broad

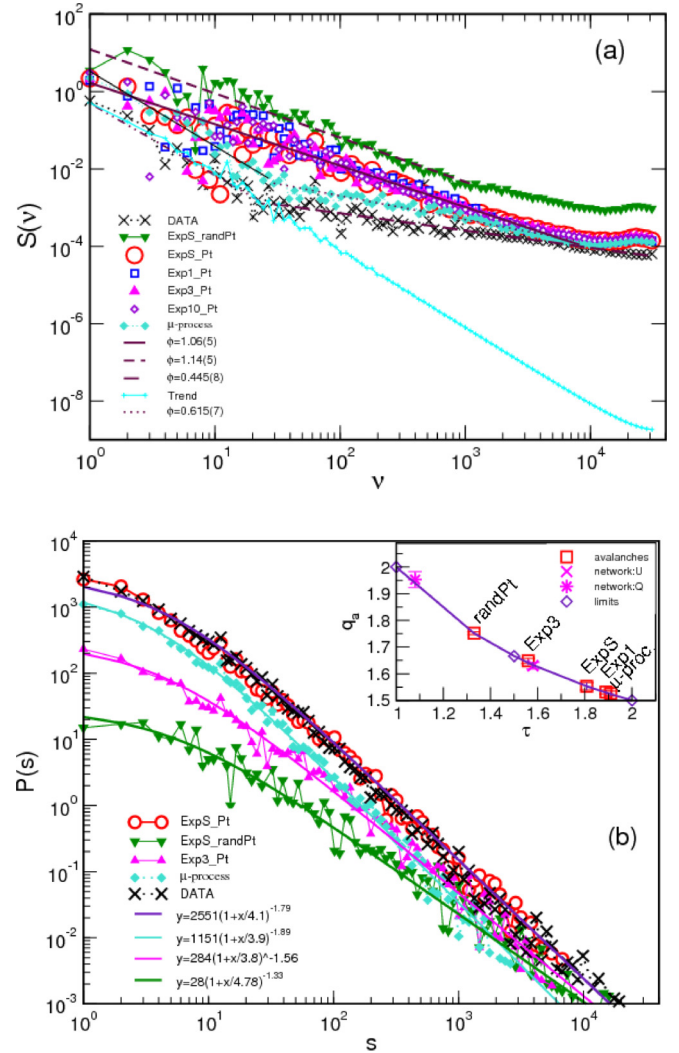


FIG. 6. The (a) power spectra and (b) avalanche size distributions in the empirical data and ABM for different expertise. Inset: The nonextensivity parameter  $q_a$  vs  $\tau_s$  for the considered cases.

range of frequencies  $\nu$ . In Fig. 6(a), we show the results for the power spectrum of the time series of events in the empirical data and the simulated signals for different agent's expertise and two driving modes. The corresponding distributions of the avalanche sizes obtained from these time series are shown in Fig. 6(b). These figures indicate that an extended scaling range occurs over several orders of magnitude in the power spectrum as well as in the avalanche sizes. However, the driving mode and the actor's expertise affect the scale invariance in a different manner. Specifically, the increasing trend in the driving signal  $p(t)$  is pronounced in the power spectrum of the empirical data and  $\mu$  process. Whereas, the increased activity in the innovation layer is balanced by the strict expertise-matching, resulting in the correlations of the flicker-noise type (middle curves). When the driving rate is elevated, i.e., *randPt* case, the slope  $\phi$  increases in the region of low frequencies and decreases in the high-frequency region (top curve).

While the shape of the driving signal is essential for the scaling in the power spectrum, the scale invariance of the avalanches is equally sensitive to the actual expertise of

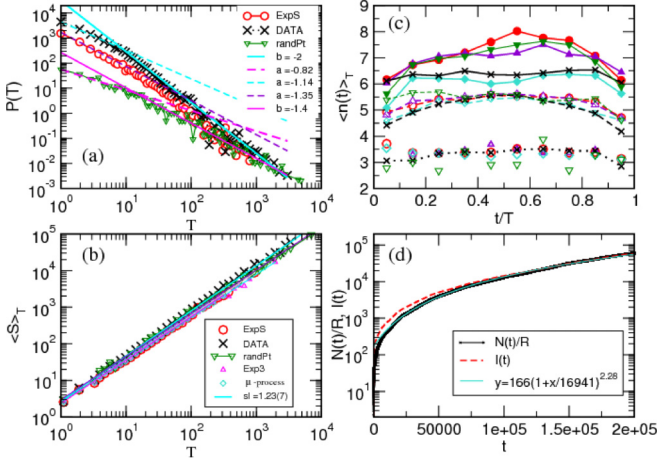


FIG. 7. (a) The distribution of avalanche duration and (b) the average size  $\langle S \rangle_T$  of the avalanches of a fixed duration  $T$  plotted against  $T$  for the empirical data and simulations with varied expertise and driving rate, as indicated. (c) The average activity  $\langle n(t) \rangle_T$  within an avalanche of the duration  $T$  averaged over the avalanches in three ranges:  $T \leq 10$  (half-filled symbols, dotted line),  $T \in (10, 100]$  (open symbols, dashed lines), and  $T > 100$  (filled symbols, full lines). The symbol shape and color correspond to the legend in the panel (b). (d) Extracted from the empirical data, the innovation  $I(t)$  increase with time and the integrated activity time series  $N(t) = \sum_i n(t)$ , normalized by the ratio  $R = \langle n(t) \rangle / \langle dI(t)/dt \rangle = 32.5$ .

the agents. The slopes of the distribution of the avalanche sizes  $\tau_s$  are found in the range from 1.33 to 1.93, depending on the expertise and driving; note that the degree distributions of user and question partitions in Fig. 3 are within this range (see also Discussion). However, the mathematical expressions that fit these distributions are different. In particular, the distribution of avalanche sizes, shown in Fig. 6(b) can be fitted by the  $q_a$ -exponential function

$$P(S) = A[1 - (1 - q_a)S/S_0]^{1/q_a}, \quad (4)$$

where the parameter  $q_a > 1$  measures the degree of nonextensivity in the underlying stochastic process [40,59–62]. The values of the scaling exponents of the avalanche size distribution and the corresponding values of the nonextensivity parameter  $q_a$  are shown in the inset in Fig. 6(b). Notably, the distribution of the avalanche sizes obtained from the empirical data and simulations with the expertise of the agents *ExpS* taken from the empirical distribution are similar and close to the case of  $\mu$  process. The probability of large avalanches increases with the increased average expertise; the case *Exp3* is shown. It is important to stress that the same type of distribution with a power-law tail is also obtained in the case of *randPt*, representing an increased driving rate with the accelerated network growth, as mentioned above. Moreover, the exponent of the distribution of avalanche sizes is smaller than when the lower driving rate pertinent to  $p(t)$  signal is used; see also Fig. 7(a) for the distribution of duration. The decrease of the scaling exponents provides an observable measure of the effects of avalanche merging, which occurs more often at the elevated driving rate, to the critical state. These findings are in agreement with the studies of SOC in cellular

automata [63] and physical systems [64] under various driving rates.

## B. Propagation and geometry of avalanches

Statistics of the avalanche sizes with power-law tails, as shown in Fig. 6(b), is compatible with the occurrence of self-organized dynamics. In the following, we show that the propagation of avalanches and their shapes further confirm these features of the underlying dynamics. The time-dependent characteristics of the avalanches evaluated for the above-studied sets of parameters are demonstrated in Figs. 7(a)–7(c). Specifically, the distribution of the duration  $T$  of avalanches, cf. Fig. 7(a), exhibits different scaling behavior for small avalanches with the duration  $T < T_x \sim 10$  as compared to the asymptotic scaling law  $P(T) \sim T^b$ . In the asymptotic region, we find that (within the numerical error bars)  $b = -2$  for the avalanche durations in the empirical data and a close value  $b = -1.96$  for the simulated system with *ExpS*. Gradually smaller slopes are found for the cases *Exp3* (not shown) and *randPt*. Similarly, the scaling exponents in the short-avalanche region vary with the expertise and the driving rate from  $a = -0.82$  for *randPt* to  $a = -1.35$  for *ExpS*. The corresponding range for the empirical data is even shorter; see Fig. 7(a). Given different shapes, the size of the avalanches of a fixed duration  $T$  can vary; cf. Fig 4(b) and Eq. (1). Nevertheless, in the SOC systems, the average size  $\langle S \rangle_T$  of the avalanches of a given duration  $T$  scales with  $T$  according to Eq. (3). Here, the tail exponent  $b$  of the duration distribution appears, i.e.,  $\langle S \rangle_T \sim T^{\gamma_{ST}}$ , where the exponent  $\gamma_{ST} = (b - 1)/(\tau_s - 1)$ . Figure 7(b) exhibits the plots  $\langle S \rangle_T$  against  $T$  corresponding to the avalanches studied in this work. The apparent power-law dependence in this plot suggests that the scale invariance of the avalanches can not be affected by a reasonable shift of the baseline, for instance, by  $2\sigma$ . The average exponent  $\gamma_{ST} = 1.23 \pm 0.07$  suggests rather narrow avalanches; apart from the duration range, which varies with the simulation parameters, the variations of the exponent are rather small.

The precise shape of the avalanche of the duration  $T$  is given by the sequence of the elementary pulses  $n(t)$  over time. In Fig. 7(c), we show the average height  $\langle n(t) \rangle_T$  belonging to the avalanche of a given duration  $T$  evaluated in bins of the reduced time  $t/T$ . Three groups of avalanches are distinguished, in particular, the small avalanches of the durations  $T \leq 10$ , medium-duration  $10 < T \leq 100$ , and large avalanches for the durations  $T > 100$ . As the Fig. 7(c) shows, the shape of the small avalanches is practically independent of the system's parameters. The same conclusion applies for the medium-size avalanches in the peak region, whereas they slightly differ in the decaying phase and even more in the raising phase. In the case of the large avalanches, however, the major differences occur in the peak phase. Moreover, the peak shifts towards later times when the total expertise is increased, or a larger driving rate applied.

In the panel (d) of Fig. 7, we show how the total innovation  $I(t)$  increases over time. The innovation, which is precisely defined as the number of unique combinations of tags, is obviously related to the activity but its fluctuations over time is a more subtle feature of the process, which depends on the available expertise of the agents, as shown in [44]. Here,



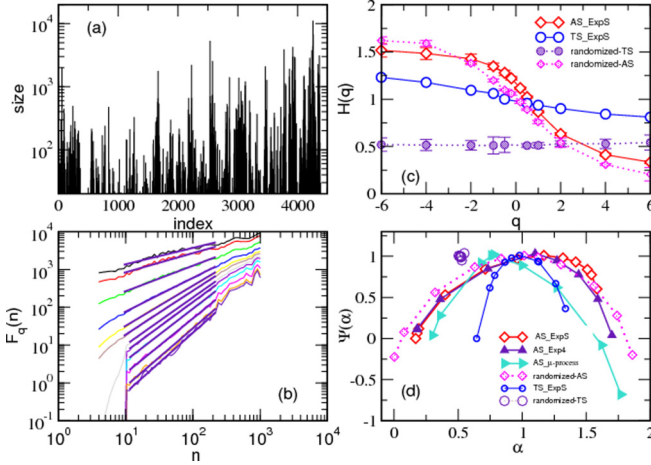


FIG. 8. For the case *ExpS*, (a) the avalanche sequence, (b) the fluctuation function of the avalanche size series, and (c) the generalized Hurst exponent  $H(q)$  for the avalanche series and the underlying time series and their randomized versions. The panel (d) shows the singularity multifractal spectrum for the avalanche sequences determined for the cases with the varied expertise and driving, as indicated in the legend.

we compute the temporal dependence of the innovation from the empirical data. The total activity (scaled by a constant factor) as a function of time is also shown in Fig. 7(d). These results suggest that, apart from the excess of the innovation in the initial period, the asymptotic law is the same for the total activity and the innovation growth with time. While the average activity is 32.5 times larger, we can conclude that the activity is driven by the fluctuations of the innovation rate. As mentioned earlier, the innovation is brought by the expertise of the actors. Hence, the increase of the innovation can be considered as a driving force for the knowledge-creation processes.

### C. Avalanche sequences and multifractality

Compared to the time series of the number of events  $n(t)$ , the avalanches are the objects occurring at a mesoscopic scale. Each avalanche consists of a certain number of the elementary pulses, cf. (1); these pulses are combined in a different way to embody the growth, peak, and relaxation phase of the avalanche. Thus, the sequence of avalanches in time contains additional information about the nature of the underlying stochastic process. For instance, the avalanche return [61,62] in many different complex systems reveals a non-Gaussian relaxation. For the knowledge-creation processes, we have demonstrated [44] that the avalanche first-return statistics obeys a  $q_r$ -Gaussian distribution with a large parameter  $q_r \approx 2.45$ . Here, we apply multifractal analysis to examine another signature of the complexity of these avalanche sequences. In particular, we analyze the temporal sequences of the avalanche sizes  $S_k$ , where  $k = 1, 2, \dots, K_{\max}$  is the index of the avalanche and  $K_{\max}$  stands for the total number of avalanches that occur in a particular time series. We consider the avalanche sequences obtained for the combinations of the parameters studied in the preceding sections and the empirical data; two examples of such avalanche sequences are displayed in Fig. 5 (bottom) and Fig. 8(a).

To determine the multifractal spectrum  $\Psi(\alpha)$  for the sequences  $S_k$ , we apply the detrended multifractal analysis (DMFA); we use the approach which was applied to different types of complex signals, as described in Refs. [65–68]. According to the standard procedure, the profile of the signal is first constructed by the integration

$$Y(i) = \sum_{k=1}^i (S_k - \langle S \rangle). \quad (5)$$

The profile is then divided into  $N_s = \text{int}(K_{\max}/n)$  nonoverlapping segments of equal length  $n$ . Then for each segment  $\mu = 1, 2, \dots, N_s$ , the local trend  $y_\mu(i)$  is determined and the standard deviation around the local trend

$$F^2(\mu, n) = \frac{1}{n} \sum_{i=1}^n \{Y[(\mu - 1)n + i] - y_\mu(i)\}^2 \quad (6)$$

is found. Similarly, the procedure is repeated starting from the end of the signal, resulting in  $F^2(\mu, n) = \frac{1}{n} \sum_{i=1}^n \{Y[N - (\mu - N_s)n + i] - y_\mu(i)\}^2$  for  $\mu = N_s + 1, \dots, 2N_s$ . Combining the deviations at all segments, the  $q$ th order fluctuation function  $F_q(n)$  is obtained according to

$$F_q(n) = \left\{ \frac{1}{2N_s} \sum_{\mu=1}^{2N_s} [F^2(\mu, n)]^{q/2} \right\}^{1/q} \sim n^{H(q)}, \quad (7)$$

and plotted against the varied segment length  $n \in [2, \text{int}(K_{\max}/4)]$ . The scale invariance of  $F_q(n)$  against the segment length  $n$  is examined to determine the corresponding scaling exponent  $H(q)$ . Here, the distortion parameter  $q$  takes a range of real values. The main idea is that the segments of the signal with potentially different fractal features will be suitably enhanced by a particular  $q$  value to become self-similar to the full signal and the corresponding scaling exponent  $H(q)$  as a function of  $q$  is measured. Notably, different *small fluctuation segments* are enhanced by the negative values of  $q$ , and the segments with *large fluctuations* dominate the fluctuation function for the positive values of  $q$ . In the limiting case of monofractal,  $H(q) = H(q = 2)$  is the standard Hurst exponent. Using the scaling relation  $\tau(q) = qH(q) - 1$ , the exponent  $\tau(q)$  of the box probability, defined in the partition function method [65], is computed. Thus, the generalized Hurst exponents  $H(q)$  can be related with the singularity multifractal spectrum via the Legendre transform  $\Psi(\alpha) = q\alpha - \tau(q)$ , where  $\alpha = d\tau/dq$  is the singularity strength.

Figures 8(a)–8(c) shows the results for the case *ExpS*, which incorporates the features of the empirical data and the expertise matching in the simulations. Specifically, the avalanche series, the corresponding fluctuation function, and the generalized Hurst exponent are shown to demonstrate the procedure. Also, we show the results of the DMFA applied to the underlying time series for the same parameters. For the comparison, the analysis is performed for the randomized signals; the corresponding scaling exponents  $H(q)$  are also depicted in Fig. 8(c). Notably, both the time series of the activity and the related avalanche series exhibit multifractal features. The span of the generalized Hurst exponent is much larger in the avalanche series. In the randomized case, the avalanche series exhibit almost unchanged multifractality while the time series of pulses becomes a monofractal with the properties of white noise [ $H(q) = 0.5$  within the numerical

error bars]. These findings indicate that the origin of the multifractality in the time series may be found in the temporal correlations, which are of the  $1/\nu$  type; see Fig. 6(a). While, for the avalanche series, the scale invariance of the distribution of sizes can be the sole reason for the observed multifractality. In this regard, it is interesting to analyze the width of the singularity spectrum  $\Psi(\alpha)$  for different parameters of the model. These results are shown in Fig. 8(d). In agreement with the values of  $H(q)$  for the case *ExpS* in the panel (c), the spectra corresponding to the avalanche series and the randomized avalanche series are wide. In contrast, the multifractality of the original time series results in the narrow range; further, the spectrum is reduced to a close vicinity of the point  $\Psi(\alpha = 0.5) = 1$ , corresponding to the monofractal randomized time series. For the varied expertise of the actors, we obtain wide spectra of the avalanche series, where the singularity strength  $\alpha \in [0.2, 1.7]$ ; cf. Fig. 8(d). Notably, the various parameters of the model mostly affect the right side of the spectrum, corresponding to  $q < 0$  region, i.e., small fluctuation segments in the avalanche series. While, the variations are less pronounced for the large fluctuation segments, appearing in the left end of the spectrum  $\Psi(\alpha)$ . In this end, the empirical data and model simulations lead to similar results. These findings indicate how the sequences of small and large avalanches are affected by the available expertise; the occurrence of the large fluctuations in the avalanche sizes might be chiefly conditioned by the number of the actors involved.

#### IV. THE IMPACT OF VACANCIES ON MULTIFRACTAL SPECTRUM: A COMPARISON TO SPIN-ALIGNMENT AVALANCHES

As mentioned in the Introduction, the complexity of the social interactions prevents the exact description of the mechanisms at the elementary scale, calling for a comparison to better-understood physics models. In this regard, the interacting spin system described by the random-field Ising model at zero temperature and driven along the hysteresis loop represents a paradigm of complex dynamical behavior far from the equilibrium [69]. In this example, the spin alignment along the slowly increasing external field is balanced by spin-spin interactions and the local constraints due to the random-field disorder. The dynamics of spin flips under the disorder induced constraints and interactions was often employed to model opinion formation [70], processes driven by social balance [71] and other cases. For the purpose of this work, we aim to explore the impact of vacancies in the underlying network onto the multifractal spectrum of the avalanche sequences. We consider spin-alignment avalanches in the zero-temperature random-field Ising model (ZTRFIMc) with the two-state spin site  $S_i(t) = \pm 1$  at each lattice site  $i = 1, 2, \dots, N$  with a fraction  $c > 0$  of defect sites where the spin is absent. The energy  $\mathcal{H} = -\sum_i \tilde{h}_i(t) S_i(t)$  is minimized by the spin alignment along the current value of the local field  $\tilde{h}_i(t)$ , where

$$\tilde{h}_i(t) = \sum_{j \in nn} J_{ij} S_j(t) + h_i + B(t); \quad S_i(t+1) = \text{sgn}[\tilde{h}_i(t)] \quad (8)$$

Here,  $h_i$  is the local quenched random field, which is described by the Gaussian distribution of zero mean and the width  $\Delta$ . Starting from a large negative value, the system is driven quasistatically, i.e., by the slowly increasing external field  $B(t+T) = B(t) + \delta B$  after an avalanche stops. The ferromagnetic interaction  $J_{ij}$  among the pair of spins at the adjacent sites  $i, j$  has the positive mean  $\langle J_{ij} \rangle = J$  and the second cumulant  $Jc(1-c)$ , where  $c > 0$  is the probability that the spin is absent at a randomly selected site. The dimensionless parameters  $f \equiv \Delta/J$  and  $r \equiv \delta B/J$  characterize the pinning strength and the driving rate, respectively.

Given the domain structure in these disordered systems, the magnetization reversal occurs in a series of jumps by the slow field ramping along the hysteresis loop. These magnetization changes are directly related to the motion of the domain walls, accompanying the expansion of the domains which are oriented parallel to the field. The size of the magnetization changes thus occur in the interplay between the driving by the external field and pinning of the domain walls at by the local random fields, oriented opposite. These magnetization changes in time represent the data points in the Barkhausen noise, a complex time series from which then the avalanches can be determined. The scale-invariant behavior of the Barkhausen avalanches and their dependence on the strength of the random-field disorder has been well understood [27,69,72–74]. Recently, it has been shown [68] that the Barkhausen noise exhibits multifractal structure. Moreover, the dynamical regime in the central part of the hysteresis loop, where large avalanches can occur for the weak disorder, has a significantly different spectrum from the dynamics at the beginning of the hysteresis loop. In contrast, the presence of hard defects in ZTRFIMc has been much less investigated. Specifically, even in the weak random-field pinning that allows system-wide avalanches, the presence of hard defects induces a characteristic length, which affects the cutoff size of the avalanches and also the universality of the scaling exponents [75,76]. Here, we are interested in the dynamics of the avalanches in the presence of site defects.

We consider a small concentration of the randomly dispersed site defects  $c = 0.05$  on top of the weak random field disorder and slow driving; thus we use a representative set of parameters in this regime [68]:  $f = 2.3 < f_c$  and  $r = 0.02$  in the three-dimensional cubic lattice of  $100^3$  spins. In the absence of the site defects ( $c = 0$ ), the system of this dimension would undergo a domain-wall depinning via large avalanches in the central part of the hysteresis loop [68]. However, the small percentage of site defects suffices to perturb this critical behavior by the pinning of the domain walls at a distance  $\ell \sim 1/c$ ; whereas, at the distances  $x \lesssim \ell$  the domain wall motion is accelerated by the external field, corresponding to the regime of the weak random-field pinning. Consequently, the small avalanches are similar as in the case of weak pinning without site defects, while the propagation of large avalanches is considerably hindered. These effects are reflected in the multifractal spectrum of the avalanche sequences, in particular, by increasing the difference in the generalized Hurst exponent for  $q > 0$  and  $q < 0$  values. The resulting singularity multifractal spectra are shown in the inset to Fig. 9. Although the avalanches tend to be larger in the central part of the hysteresis loop, the presence of site defects

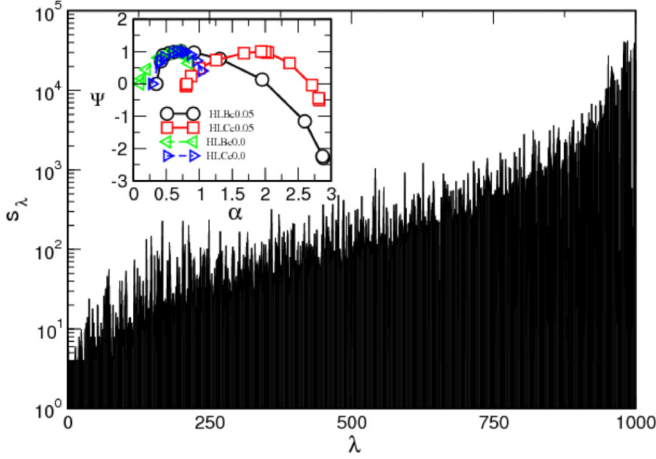


FIG. 9. The avalanche sequence at the beginning of the hysteresis loop (HLB) of the ZTRFIMc for  $c = 0.05$  and weak random field pinning. Inset: The singularity multifractal spectra in the hysteresis loop center (HLC) and the loop beginning (HLB) for  $c = 0.05$  and  $c = 0.0$ .

induces a mixture of small and large events, which results in the substantial broadening of the spectrum in comparison to the case of  $c = 0$ .

It is interesting to point the similarity between the spectra in the avalanche series that are affected by site defects in Fig. 9 with the spectra of social avalanches of knowledge creation, Fig. 8. In this case, the avalanche propagation is conditioned by the actor's activity patterns and the (non)possession of the required expertise. According to Fig. 2(a), a large number of users (agents) stops to be active after a given number of actions, while their artifacts are still available and can be the subject of interest to others. Thus, each avalanche becomes pinned by reaching the network node of such a user, in the manner that a site defect pins the moving domain wall. Also, the users whose delay time is typically larger than the avalanche duration may have a similar effect on the current avalanche propagation. On the other hand, the extremely active users [a small fraction, represented by the end of the distribution in Fig. 2(a)] accelerate the propagation through their numerous connections and also by being active at more than one question within a short time interval. This type of the actor's heterogeneity results in a typical mixture of small and large events, as seen in the analysis in Sec. III. Beyond the shape of the multifractal spectra, representing a combination of small and large events, the avalanche distributions differ in the random-site ferromagnetic model. Among other reasons, the evolving bipartite networks of the actors and their artifacts are identified as those being of the key importance.

## V. DISCUSSION AND CONCLUSIONS

Considering the large data set from Q&A site Mathematics Stack Exchange and the agent-directed modeling, we have analyzed the avalanching behavior and the bipartite network that underlies the creation of collective knowledge. Given the complexity of modeling the human actors, we have kept the agent's properties statistically similar to the features of the users, detectable from the same empirical data. In

particular, the agent's activity pattern is designed by the distributions of the number of actions per user  $P(N_i)$  and the interactivity time  $P(\Delta t)$  as well as the arrival rate  $p(t)$ , which are mutually interconnected and characterize the human dynamics in the considered empirical system. We have varied the agent's expertise, as the most relevant feature to the knowledge creation. To evaluate its impact, the expertise matching to the contents of the artifact has been clearly observed in the simulations. (Other potential extensions of the model, e.g., neglecting or altering the above distributions independently, that goes beyond human dynamics [50], are not discussed in this work.)

The analysis indicates that the knowledge-creation process represents a particular class of social dynamics, in which the self-tuning towards the criticality is controlled by the use of knowledge in the *meaningful* actions and the coevolution of the underlying network. Our key findings are here discussed.

*The self-organized criticality.* The robustly observed temporal correlations, avalanching and multifractality, as well as the scale-invariance dependence on the driving rate, indicate that the criticality might occur in these stochastic processes in a self-tuned manner. The social dynamics, driven by the arrival of new actors and innovation that they bring, represents the main source of the avalanching behaviors. Whereas, the altered degree of branching (e.g., in the  $\mu$  process) and, more importantly, the strict use of the expertise imposes the constraints to the social dynamics, which affects the avalanche propagation. These constraints then manifest in the nonuniversal scaling exponents of the avalanche sizes and durations. Moreover, the relative fraction of the small and large avalanches, which appear to be mixed in the course of the process, depends on the available expertise of the actors, thus affecting the width of the multifractal spectrum. Nonextensivity ( $q_a > 1$ ) is another remarkable feature of the knowledge-creation process, where the  $q_a$ -exponential distribution applies to the avalanche sizes and the  $q_r$ -Gaussian distribution to the avalanche returns. Although the observed multifractality of the avalanche series is compatible with these distributions, more theoretical work is needed to unravel their origin.

*The structure of network partitions.* In this process, the growth of the bipartite network occurs by the addition of layers, which contain new arrivals and other active agents, and their artifacts. How the network will grow is strongly related to the expertise-conditioned linking. After a sufficient time, the network exhibits a broad distribution of the degree of nodes in both partitions. The scaling exponent of the agent's degree distribution in each case is close to the introduced distribution of the activity  $P(N_i)$ . However, the cutoffs of the distributions, indicating the actual size of the network, vary with the considered apportionment of the expertise. Notably, the network size increases when the agents possess a larger expertise in the average. In this case, the probability of an agent to connect to a suitable artifact is elevated. Similarly, the network growth is accelerated by the addition of a larger number of actors in the initial stages of the process, which results in a greater number of the available artifacts. Thus, in the social process, the scale freeness of the user partition is determined by the actual activity profiles alone. Where the corresponding edges will appear in the network, and,

consequently the network's community structure [44], depends on the available expertise and how it is distributed over the actors. On the other hand, the question-partition degree distribution strongly depends on the expertise. Note that in both cases the determined scaling exponents are smaller than 2. These results suggest that the effective mechanisms might be different from the popular "preferential" attachment-andrewiring rule that leads to the exponents larger than 2.

*The self-tuning factors.* The heterogeneity of the actor's profiles and the activity patterns, with the extended range of delay times and the number of actions, is certainly an important factor for the appearance of the avalanches in the studied social dynamics. Apart from these "human factors," it is stipulated that some other ingredients of the knowledge creation may contribute to the paths towards the criticality. Specifically, the importance and the use of expertise in the human collaborative endeavor makes the social process substantially different from, for instance, the activity on popular blogs, where the negative emotional charge of comments may lead to a supercritical avalanche [13,14]. Here, the required expertise of the actors needs to match the contents of the question, resulting in a balanced activity (resembling the energy balance in the driven physical system). Consequently, the activity stops when sufficient knowledge is built through the answers on a particular question, depending on the available expertise of the actors. The same artifacts may become a focus in the later stages when new necessary knowledge becomes available, i.e., by the arrival of new players. Thus, in contrast to the standard social dynamics, a kind of optimization of the available expertise applies, which is also compatible with the nonextensive dynamics mentioned above. Note that the optimization of the system's efficiency is often associated with the functioning of biological systems [7] and the avalanching process in neuronal assemblies [8,9] and the brain [12], which are still not well understood. Furthermore, the underlying network evolution by the addition of the innovative contents in the active layer can be seen as another decisive factor to provide a particular type of critical behavior. Theoretically, changing the random environment for the self-organizing process affects the universality of the critical behavior that can be achieved. The renormalization group study of the critical sandpile model in the presence of quenched [43] or annealed [42] random currents has demonstrated that a new stable fixed point appears, which is controlled by the variance of the random variable. For the knowledge creation in the online Q&A communities, it is relevant to mention that the innovation expansion builds

the network of contents with a logical structure, which originates in the participant's individual knowledge. For the same empirical data, this aspect of knowledge creation was demonstrated in [47] by analysis of the knowledge network (containing the cognitive contents that are used in all questions and answers, and encoded by the standard mathematical classification scheme). Further research is needed to disclose the potential importance of avalanches of knowledge creation in the off-line social communities, where knowledge sharing can lead to the creation of a common opinion and other collective behavior [13,19].

In summary, the creation of collective knowledge through questions and answers is a self-organized critical process where the mechanisms of self-tuning are provided by the interplay of the social and cognitive layer. The observed SOC is robust to the increased driving rate within limits pertinent to the considered experimental system. Our study suggests some questions for further theoretical considerations, in particular: the formal differences between knowledge creation and common social dynamics; the potential similarity between the knowledge creation and the brain avalanching dynamics; the origin of the nonextensivity (although the nonextensive character of the dynamics is intuitive in the context of knowledge, the formal origin of the  $q$ -Gaussian fluctuations is not understood), and other issues. The presented results, based on the empirical data and the agent-directed model, which is almost equally complex as the empirical system itself, reveal many factors that act in unison and contribute to the observed SOC. The presented comparison to the driven spin system with site defects suggests that apart from the expertise, the heterogeneity of the actor's activity patterns is an essential factor that prevents the appearance of the supercritical avalanches. Our findings may help design formal theoretical models of SOC, e.g., of the cellular automata type or the continuous models suitable for the renormalization group analysis, which may be capable of describing the unique role of each of these factors.

#### ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Slovenian Research Agency under the Program P1-0044 and the Ministry of Education, Science and Technological Development of the Republic of Serbia under the Project ON171017.

- 
- [1] D. Marković and C. Gros, Power laws and self-organized criticality in theory and nature, *Phys. Rep.* **536**, 41 (2014).
  - [2] M. J. Aschwanden, *Self-organized criticality systems* (Open Academic Press, Berlin, 2013).
  - [3] *Avalanches in Functional Materials and Geophysics*, edited by E. K. H. Salje, A. Saxena, and A. Planes (Springer International Publishing AG, Cham, Switzerland, 2017).
  - [4] P. Bak, C. Tang, and K. Wiesenfeld, Self-Organized Criticality: An Explanation of the  $1/f$  Noise, *Phys. Rev. Lett.* **59**, 381 (1987).
  - [5] D. Dhar, Self-Organized Critical State of Sandpile Automaton Models, *Phys. Rev. Lett.* **64**, 1613 (1990).
  - [6] B. Tadić and D. Dhar, Emergent Spatial Structures in Critical Sandpiles, *Phys. Rev. Lett.* **79**, 1519 (1997).
  - [7] T. Mora and W. Bialek, Are biological systems poised at criticality? *J. Stat. Phys.* **144**, 268 (2011).
  - [8] A. Levina, J. M. Herrmann, and T. Geisel, Dynamical synapses causing self-organized criticality in neural networks, *Nat. Phys.* **3**, 857 (2007).

- [9] J. G. Orlandi, J. Soriano, E. Alvarez-Lacalle, S. Teller, and J. Casademunt, Noise focusing and the emergence of coherent activity in neuronal cultures, *Nat. Phys.* **9**, 582 (2013).
- [10] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Scale-free correlations in starling flocks, *Proc. Natl. Acad. Sci. USA* **107**, 11865 (2010).
- [11] D. S. Zhukov, V. V. Kanishchev, and S. K. Lyamin, Application of the theory of self-organized criticality to the investigation of historical processes, *SAGE Open* **6**, 2158244016683216 (2016).
- [12] J. Hesse and T. Gross, Self-organized criticality as a fundamental property of neural systems, *Front. Syst. Neurosci.* **8**, 166 (2014).
- [13] B. Tadić, V. Gligorijević, M. Mitrović, and M. Šuvakov, Co-evolutionary mechanisms of emotional bursts in online social dynamics and networks, *Entropy* **15**, 5084 (2013).
- [14] M. Mitrović and B. Tadić, Bloggers behavior and emergent communities in blog space, *Eur. Phys. J. B* **73**, 293 (2010).
- [15] M. Mitrović, G. Paltoglou, and B. Tadić, Quantitative analysis of bloggers' collective behavior powered by emotions, *J. Stat. Mech.* (2011) P02005.
- [16] M. Šuvakov, M. Mitrović, V. Gligorijević, and B. Tadić, How the online social networks are used: Dialogues-based structure of MySpace, *J. R. Soc., Interface* **10**, 20120819 (2012).
- [17] M. Šuvakov and B. Tadić, Collective emotion dynamics in chats with agents, moderators and bots, *Condens. Matter Phys.* **17**, 33801 (2014).
- [18] Yu. Holovatch, O. Mrygold, M. Szell, and S. Thurner, Analysis of a virtual world, in *Math Meets Myths: Quantitative Approaches to Ancient Narratives*, edited by R. Kenna *et al.*, Understanding Complex Systems (Springer International Publishing Switzerland, 2017).
- [19] G. Kou, Y. Zhao, Y. Peng, and Y. Shi, Multi-level opinion dynamics under bounded confidence, *PLoS One* **7**, e43507 (2012).
- [20] M. Malkov, A. Das, I. Gruzinov, M. N. Rosenbluth, C. Holland, B. Wecht, A. I. Smolyakov, F. L. Hinton, Z. Lin, T. S. Hahm, P. H. Diamond, and S. Champeaux, Secondary instability in drift wave turbulence as a mechanism for zonal flow and avalanche formation, *Nucl. Fusion* **41**, 1067 (2001).
- [21] Y. Idomura, H. Urano, N. Aiba, and S. Tokuda, Study of ion turbulent transport and profile formations using global gyrokinetic full-f Vlasov simulation, *Nucl. Fusion* **49**, 065029 (2009).
- [22] K. P. Zybin and V. A. Sirota, Multifractal structure of fully developed turbulence, *Phys. Rev. E* **88**, 043017 (2013).
- [23] Á. Corral, Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes, *Phys. Rev. Lett.* **92**, 108501 (2004).
- [24] C. Narteau, S. Byrdina, P. Shebalin, and D. Schorlemmer, Common dependence on stress for the two fundamental laws of statistical seismology, *Nature (London)* **462**, 642 (2009).
- [25] F. Dalton and D. Corcoran, Basin of attraction of a bounded self-organized critical state, *Phys. Rev. E* **65**, 031310 (2002).
- [26] T. Hatano, C. Narteau, and P. Shebalin, Common dependence on stress for the statistics of granular avalanches and earthquakes, *Sci. Rep.* **5**, 12280 (2015).
- [27] D. Spasojević, S. Bukvić, S. Milošević, and H. E. Stanley, Barkhausen noise: Elementary signals, power laws, and scaling relations, *Phys. Rev. E* **54**, 2531 (1996).
- [28] A. Berger, A. Inomata, J. S. Jiang, J. E. Pearson, and S. D. Bader, Experimental Observation of Disorder-Driven Hysteresis-Loop Criticality, *Phys. Rev. Lett.* **85**, 4176 (2000).
- [29] H.-S. Lee, K.-S. Ryu, I.-S. Kang, and S.-C. Shin, Universal Barkhausen critical scaling behavior observed in  $\text{Ni}_x\text{Fe}_{1-x}$  ( $x = 0-0.5$ ) films, *J. Appl. Phys.* **109**, 07E101 (2011).
- [30] L. Carrillo and J. Ortín, Avalanches in the growth of stress-induced martensites, *Phys. Rev. B* **56**, 11508 (1997).
- [31] D. Soto-Parra, X. Zhang, S. Cao, E. Vives, E. K. H. Salje, and A. Planes, Avalanches in compressed Ti-Ni shape-memory porous alloys: An acoustic emission study, *Phys. Rev. E* **91**, 060401 (2015).
- [32] G. F. Nataf, P. O. Castillo-Villa, J. Baró, X. Illa, E. Vives, A. Planes, and E. K. H. Salje, Avalanches in compressed porous  $\text{SiO}_2$ -based materials, *Phys. Rev. E* **90**, 022405 (2014).
- [33] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. Lee DeVillie, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Universal Critical Dynamics in High Resolution Neuronal Avalanche Data, *Phys. Rev. Lett.* **108**, 208102 (2012).
- [34] M. Šuvakov and B. Tadić, Modeling collective charge transport in nanoparticle assemblies, *J. Phys.: Condens. Matter* **22**, 163201 (2010).
- [35] E. Altshuler, T. H. Johansen, Y. Paltiel, P. Jin, K. E. Bassler, O. Ramos, Q. Y. Chen, G. F. Reiter, E. Zeldov, and C. W. Chu, Vortex avalanches with robust statistics observed in superconducting niobium, *Phys. Rev. B* **70**, 140505 (2004).
- [36] M.-Y. Im, K.-S. Lee, A. Vogel, J.-I. Hong, G. Meier, and P. Fischer, Stochastic formation of magnetic vortex structures in asymmetric disks triggered by chaotic dynamics, *Nat. Commun.* **5**, 5620 (2014).
- [37] X. Jiang, L. Thomas, R. Moriya, M. Hayashi, B. Bergman, C. Rettner, and S. S. P. Parkin, Enhanced stochasticity of domain wall motion in magnetic racetracks due to dynamic pinning, *Nat. Commun.* **1**, 25 (2010).
- [38] V. D. Nguyen, W. Saverio Torres, P. Laczkowski, A. Marty, M. Jamet, C. Beigné, L. Notin, L. Vila, and J. P. Attané, Elementary depinning processes of magnetic domain walls under fields and currents, *Sci. Rep.* **4**, 6509 (2014).
- [39] B. Tadić, Dynamic criticality in driven disordered systems: Role of depinning and driving rate in Barkhausen noise, *Physica A* **270**, 125 (1999).
- [40] *Nonextensive Entropy—Interdisciplinary Applications*, edited by M. Gell-Mann and C. Tsallis, Santa Fe Institute Studies in the Science of Complexity (Oxford University Press, Oxford, 2004).
- [41] U. C. Täuber, Phase transitions and scaling in systems far from equilibrium, *Annu. Rev. Condens. Matter Phys.* **8**, 185 (2017).
- [42] N. V. Antonov, and P. I. Kakin, Effects of random environment on a self-organized critical system: Renormalization group analysis of a continuous model, *EPJ Web Conf.* **108**, 02009 (2016).
- [43] B. Tadić, Disorder-induced critical behavior in driven diffusive systems, *Phys. Rev. E* **58**, 168 (1998).
- [44] M. M. Dankulov, R. Melnik, and B. Tadić, The dynamics of meaningful social interactions and the emergence of collective knowledge, *Sci. Rep.* **5**, 12197 (2015).
- [45] J. Kimmerle, U. Kress, and C. Held, The interplay between individual and collective knowledge: technologies for organisational learning, *Knowledge Management Research Practice* **8**, 33 (2010).

- [46] *Social Interactions and the Development of Knowledge*, edited by J. I. M. Carpendale and U. Müller (Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2013).
- [47] M. Andjelković, B. Tadić, M. M. Dankulov, M. Rajković, and R. Melnik, Topology of innovation spaces in the knowledge networks emerging through questions-and-answers, *PLoS ONE* **11**, e0154655 (2016).
- [48] M. Mitrović, G. Paltoglou, and B. Tadić, Networks and emotion-driven user communities at popular blogs, *Eur. Phys. J. B* **77**, 597 (2010).
- [49] M. Mitrović and B. Tadić, Dynamics of bloggers' communities: Bipartite networks from empirical data and agent-based modeling, *Physica A* **391**, 5264 (2012).
- [50] B. Tadić, Modeling behavior of Web users as agents with reason and sentiment, in *Advances in Computational Modeling Research: Theory, Developments and Applications*, edited by A. B. Kora (Nova Publishing, New York, 2013), pp. 177–186.
- [51] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [52] *Complex Networks and their Applications*, edited by H. Cherifi (Cambridge Scholar Publishing, Newcastle upon Tyne, 2014).
- [53] T. Nishikawa, J. Sun, and A. E. Motter, Sensitive dependence of network dynamics on network structure, [arXiv:1611.01164](https://arxiv.org/abs/1611.01164).
- [54] B. Tadić and S. Thurner, Information super-diffusion on structured networks, *Physica A* **332**, 566 (2004).
- [55] N. E. Savitskaya, Stabilization of avalanche processes on dynamical networks, *JETP Lett.* **103**, 206 (2016).
- [56] P. Fronczak, A. Fronczak, and J. A. Hołyst, Self-organized criticality and coevolution of network structure and dynamics, *Phys. Rev. E* **73**, 046117 (2006).
- [57] D. J. B. Soares, C. Tsallis, A. M. Mariz, and L. R. Da Silva, Preferential attachment growth model and nonextensive statistical mechanics, *Europhys. Lett.* **70**, 70 (2005).
- [58] L. Hérbert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé, Constrained growth of complex scale-independent systems, *Phys. Rev. E* **93**, 032304 (2016).
- [59] C. Tsallis and D. J. Bukman, Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostistical basis, *Phys. Rev. E* **54**, R2197 (1996).
- [60] C. Tsallis, The nonadditive entropy  $S_q$  and its applications in physics and elsewhere: Some remarks, *Entropy* **13**, 1765 (2011).
- [61] G. P. Pavlos, L. P. Karakatsanis, M. N. Xenakis, E. G. Pavlos, A. C. Iliopoulos, and D. V. Sarafopoulos, Universality of non-extensive Tsallis statistics and time series analysis: Theory and applications, *Physica A* **395**, 58 (2014).
- [62] F. Caruso, A. Pluchino, V. Latora, S. Vinciguerra, and A. Rapisarda, Analysis of self-organized criticality in the Olami-Feder-Christensen model and in real earthquakes, *Phys. Rev. E* **75**, 055101 (2007).
- [63] Á. Corral and M. Paczuski, Avalanche Merging and Continuous Flow in a Sandpile Model, *Phys. Rev. Lett.* **83**, 572 (1999).
- [64] F. J. Pérez-Reche, B. Tadić, L. Mañosa, A. Planes, and E. Vives, Driving Rate Effects in Avalanche-Mediated First-Order Phase Transitions, *Phys. Rev. Lett.* **93**, 195701 (2004).
- [65] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, *Physica A* **316**, 87 (2002).
- [66] A. N. Pavlov and V. S. Anishchenko, Multifractal analysis of complex signals, *Phys. Usp.* **50**, 819 (2007).
- [67] M. S. Movahed, G. R. Jafari, F. Ghasemi, S. Rahvar, and M. R. R. Tabar, Multifractal detrended fluctuation analysis of sunspot time series, *J. Stat. Mech.* (2006) P02003.
- [68] B. Tadić, Multifractal analysis of Barkhausen noise reveals the dynamic nature of criticality at hysteresis loop, *J. Stat. Mech.* (2016) 063305.
- [69] T. P. Handford, F. J. Perez-Reche, and S. Taraskin, Mechanisms of evolution of avalanches in regular graphs, *Phys. Rev. E* **87**, 062122 (2013).
- [70] *Biology, Sociology, Geology by Computational Physicists*, edited by D. Stauffer, S. Moss De Oliveira, P. de Oliveira, and J. de Sá Martins (Elsevier, Amsterdam, 2006), Vol. 1.
- [71] Q.-K. Meng, Self-organized criticality in small-world networks based on the social balance dynamics, *Chin. Phys. Lett.* **28**, 118901 (2011).
- [72] F. J. Pérez-Reche and E. Vives, Finite-size scaling analysis of the avalanches in the three-dimensional Gaussian random-field Ising model with metastable dynamics, *Phys. Rev. B* **67**, 134421 (2003).
- [73] F. J. Pérez-Reche and E. Vives, Spanning avalanches in the three-dimensional Gaussian random-field Ising model with metastable dynamics: Field dependence and geometrical properties, *Phys. Rev. B* **70**, 214422 (2004).
- [74] F. J. Perez-Reche, C. Triguero, G. Zanzotto, and L. Truskinovsky, Origin of scale-free intermittency in structural first-order phase transitions, *Phys. Rev. B* **94**, 144102 (2016).
- [75] B. Tadić, Nonuniversal Scaling Behavior of Barkhausen Noise, *Phys. Rev. Lett.* **77**, 3843 (1996).
- [76] B. Tadić, Switching current noise and relaxation of ferroelectric domains, *Eur. Phys. J. B* **28**, 81 (2002).