Physica A 391 (2012) 5264–5278

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Dynamics of bloggers' communities: Bipartite networks from empirical data and agent-based modeling

Marija Mitrović, Bosiljka Tadić *

Department of Theoretical Physics, Jožef Stefan Institute, Box 3000 SI-1001 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 10 April 2012
Received in revised form 1 June 2012
Available online 13 June 2012

Keywords:
Bipartite networks
Online social dynamics
Agent-based modeling

ABSTRACT

We present an analysis of the empirical data and the agent-based modeling of the emotional behavior of users on the Web portals where the user interaction is mediated by posted comments, like Blogs and Diggs. We consider the dataset of discussion-driven popular Diggs, in which all comments are screened by machine-learning emotion detection in the text, to determine positive and negative valence (attractiveness and aversiveness) of each comment. By mapping the data onto a suitable bipartite network, we perform an analysis of the network topology and the related time-series of the emotional comments. The agent-based model is then introduced to simulate the dynamics and to capture the emergence of the emotional behaviors and communities. The agents are linked to posts on a bipartite network, whose structure evolves through their actions on the posts. The emotional states (arousal and valence) of each agent fluctuate in time, subject to the current contents of the posts to which the agent is exposed. By an agent's action on a post its current emotions are transferred to the post. The model rules and the key parameters are inferred from the considered empirical data to ensure their realistic values and mutual consistency. The model assumes that the emotional arousal over posts drives the agent's action. The simulations are preformed for the case of constant flux of agents and the results are analyzed in full analogy with the empirical data. The main conclusions are that the emotion-driven dynamics leads to long-range temporal correlations and emergent networks with community structure, that are comparable with the ones in the empirical system of popular posts. In view of pure emotion-driven agents actions, this type of comparisons provide a quantitative measure for the role of emotions in the dynamics on real blogs. Furthermore, the model reveals the underlying mechanisms which relate the post popularity with the emotion dynamics and the prevalence of negative emotions (critique). We also demonstrate how the community structure is tuned by varying a relevant parameter in the model. *All data used in these works are fully anonymized.*

1. Introduction

The importance of Blogs in overall information and knowledge landscape has been pointed out recently in Ref. [1], as speedy routes to promote information and get public feedback. Blogs draw their popularity from specific features: fast communication, wide accessibility, and virtual absence of editorial control. Hence, the massive use of Blogs results in a large amount of data and, similarly to other online communication media, can lead to new collective phenomena among the users [2–14]. This places new challenges for both science and the practical use of Blogs: (i) how to quantitatively measure the

* Corresponding author. Tel.: +386 14773767; fax: +386 14773724.
E-mail address: bosiljka.tadic@ijs.si (B. Tadić).
URL: http://www-f1.ijs.si/~tadic/ (B. Tadić).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2012.06.004

impact of blogging, and (ii) how to understand and potentially influence the collective phenomena through the behaviors of individual bloggers. In this work we attempt to study these aspects of blogging dynamics by using analysis of the empirical data and theoretical modeling.

Physics of complex systems and, in particular, the statistical physics of social dynamics, are focused on the dynamical processes in which human collective behaviors emerge from large number of individual actions [5,6,8]. Combining the concepts of statistical physics with the machine-learning methods for the emotion detection in texts of messages [15,16], we have recently performed an analysis of large datasets from BBC *blog.com* and *dig.com* portals and determined quantitative measures of the collective behaviors in which the emotions are involved [12,13]. Complementary to our work in Refs. [12,13], where the empirical data are analyzed to extract various complex-systems properties, the present work extends the study of the blogging dynamics in two ways. We perform the data-driven and theoretical analysis of the processes, underlying the emergence of the collective emotional behavior of Blog users, within the framework of the agent-based modeling.

The quantitative analysis of users collective behavior in the empirical data from *dig.com* and *BBC blog.com* in Refs. [12,13] is enabled by mapping the high-resolution data onto *bipartite networks* consisting of users and posts, as two natural partitions. The idea of bipartite networks also makes the "firm ground" in the present theoretical model, where the *agents interact indirectly over the posts*. In addition, we make use of several other features, observed in various empirical data, that are relevant for designing the dynamic rules of the theoretical model, specifically:

- *Universality of user's behavior* related with the action-delay and the presence of the circadian cycles [8,9,6];
- *User communities* occurring in cyberspace are reminiscent to the ones in real life, however, different time scales and grouping mechanisms might be involved [9,12,13];
- *Quantitative measures of emotions* have been introduced in psychology research [17]. In particular, based on Russell's multidimensional model of affect [18], each known emotion can be represented by a set of numerical values in the corresponding multidimensional space. Two fundamental components of emotion, to which we refer in this work, are the arousal, related to reactivity to a stimulation, and the valence, measuring intrinsic attractiveness or aversiveness to a stimulation. These components of emotion can be measured in a laboratory based on the related psycho–physiological and neurological activity [19,20]. Moreover, a systematic association has been recognized [21] between individual emotional characteristics and word use. The arousal and the valence components of an emotion can be retrieved from a written text by suitable machine-learning methods, which are being developed for a specific type of data [20,22,16].

Systematic analysis of the patterns of user behaviors and the emotional contents in the texts of comments in the empirical data from popular Blogs [12] and discussion-driven Diggs [13], suggests that negative emotions (critique) drive the activity on these Web portals. However, the mechanisms working behind this global picture have not been well understood. In order to elucidate the role of emotions in bloggers interactions, and to point out potential parameters and levels where the process can be controlled, we develop an agent-based model. The agents are communicating their emotions in a bipartite network environment. The agent's properties, the rules and the parameters of the model are derived from analysis of the empirical data of Blogs and Diggs.

Agent-based modeling [5,23,24], where different properties of agents influence their actions, provides a suitable theoretical framework for numerical simulations of social phenomena. Recently a model for product-review with the emotional agents in a mean-field environment has been introduced [24], with the agents' emotional states described by two state variables. These variables correspond to the psychological values of the arousal and the valence, respectively, in view of Russell's two dimensional circumplex model [17,18,25].

In this work we use the agent-based-modeling to explore the emergence of user communities on Blogs, where the emotional contents are communicated *indirectly via comments* that they leave on the posts. For this purpose, our emotional agents are situated on a *weighted bipartite network*, which consists of the agents (representing the users) and the posts. The bipartite network appropriately represents the actual situation on Blogs, where by definition, no link between the nodes of the same partition (i.e., agent-to-agent or post-to-post) is allowed. The weighted links between the nodes of different partitions represent the number of comments of an agent to the linked post. Motivated by the realistic situation on Blogs, in our model the network itself evolves over time due to the arrival of new agents and the addition of new posts, and due to agent's actions on previous posts of their preference. The emotional state, measured by the arousal and the valence variables, which are attached to each agent-node, is influenced by time-varying fields, stemming from the posts to which that agent is linked on the evolving network. We assume that, in analogy to real-life events, the *elevated arousal may induce an action of the emotional agent on a post*, according to the rules introduced below in Section 3. In the moment of action on a post, the agent's current emotional arousal and valence components are transferred to the comment that the agent leaves on that post, where it can be experienced by other agents. Thus the fields themselves fluctuate with the network evolution and are different for each agent, depending on its position on the network. In order to have realistic dynamics, we design the rules of actions that are motivated by systematic observations of the activity patterns in the empirical data at Blogs and Diggs, as described below in Section 2. Moreover, the set of parameters that control the dynamics of our model are inferred from the empirical data of popular discussion-driven Diggs 2.

One of the main objectives of the present work is to analyze the bipartite networks that evolve in the agent-based model and elucidate the conditions for the user communities to form within the emotion-driven dynamics. By comparing the results of the model with the corresponding ones computed from the empirical dataset, one can estimate the role of emotions in the observed users behavior in the related systems. In the following two sections we provide such comparisons

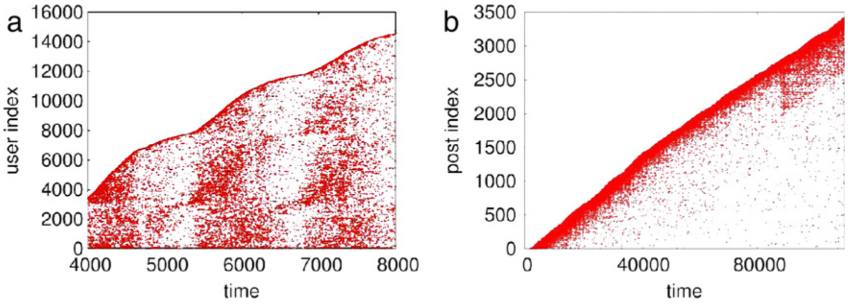

Fig. 1. Example of temporal patterns of (a) user actions and (b) activity at posts, obtained from the original dataset of discussion-driven Diggs (ddDiggs). Indexes are ordered by the user (post) first appearance in the dataset. The time is given in minutes (the original data resolution).

by computing several topology and other measures from the empirical data, in Section 2, and from the simulations, in Section 3. For reasonable comparisons, apart from the model rules the realistic values of the parameters are crucial. In Section 2 we define several quantities which characterize the empirical system and compute them from the considered dataset. These quantities are then used as the input parameters in the model. We also describe the methodology of how such parameters can be computed from any high-resolution dataset, e.g., collected from another Blog site. The remaining parameters that can not be extracted from the available datasets are considered as free parameters in the model. In Section 4 we study the effects of varying some of these parameters on the emergent network structure and the related time series. Section 5 contains a brief summary of the results and discussion.

2. Blogging dynamics and emergent networks from empirical data

The datasets that we use are collected from *BBC blog.com* and from *dig.com*. The collection methods and the structure of the data are described in detail in our previous works [12,13]. For the purpose of this paper it is important to stress that the considered data have high temporal resolution, information about identity (unique ID) of each user and of each post, and the precise relationship between the users and their comments-on-posts, as well as full text of all posts and comments. In addition, the data from *dig.com* contain information about comment-on-comment. Such information is not available in the case of BBC blogs data. Therefore, here we mostly focus on a set of the data from *dig.com*, as a good candidate where we can study the emergence of collective behavior due to the comment-mediated interaction among users [12,13].

Specifically, we select the subsets of the data which contain information related with the *popular posts*, i.e., having more than 100 comments per post, together with all users linked to them. Among these popular posts we select the subset, termed the *discussion-driven Diggs (ddDiggs)*, on which more than 50% of comments represent reply to the comments made by other users. This data consists of $N_P = 3984$ discussion-driven Digg stories, on which $N_C = 917708$ comments are written by $N_U = 82201$ users [13]. In addition, texts of posts and comments are classified by machine-learning methods with the *emotion classifier*, which is designed in Refs. [16,15] and trained at Blog-type texts. Using this emotion classifier, one can determine one component of the emotion, the valence, in the text of each comment. In the result, all comments in the considered dataset are designated as carrying either positive or negative emotion valence, or otherwise are neutral [12]. Note that in our network analysis, the emotional content of the comments is considered as an additional *property* of the post node to which the comment is addressed, and/or the link between the user and the post, along which the comment is communicated.

In the analysis of the ddDiggs dataset here we focus on the following features, which are closely related with the agent-based modeling:

- We study the temporal patterns of events, that motivate the dynamic rules of our model, and
- We define several quantities and extract their realistic values from the dataset, which are then used as the parameters in the model.

As it will be clear below, apart from its relevance for our agent-based model in Section 3, the present analysis aims for new features of these empirical datasets, which were not studied before. Specifically, this analysis yields several new results, presented in Sections 2.1 and 2.3, complementing our previous study of this data [12,13].

2.1. Extracting regularities in user behavior on popular posts

From the discussion-driven Diggs dataset, here we analyze the temporal patterns of activity related to both users and posts. Parts of these patterns are shown in Fig. 1a, b. Each user (post) occurring in that dataset is given a unique index, plotted along the vertical axis, sorted by the time of its first appearance in the dataset. For each user index, points along the time axis indicate the times when an activity of that user occurred to any one of the posts. Analogously, the points on the posts' pattern indicate the times when an activity occurred at that post by any one of the users. Although the users and the posts

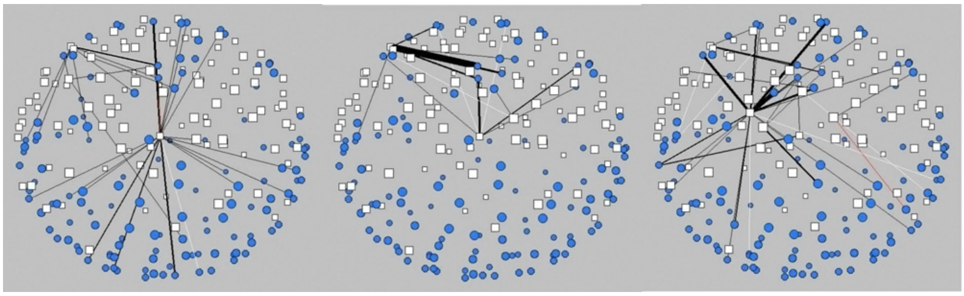

Fig. 2. Snapshots of the connections occurring within three consecutive days between the active users (•) and the active posts (□) of a small community identified on the discussion-driven Diggs. On the links, the widths represent the number of comments, and overall emotion valence of the comments is indicated by color: red—positive, black—negative, and white—or neutral. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are interconnected on the bipartite network, see part of such networks in Fig. 2, their activity patterns are entirely different. In the post-activity pattern, shown in Fig. 1b, dense points in a narrow time window following the post appearance time indicate an intensive activity at that post. This might be related with a certain *exposure of the posts* to users during that time period. The width of the exposure time window, T_0, will be recognized as a relevant parameter in the dynamics. Whereas, a different type of the dynamics beyond the exposure window is manifested in systematically reduced activity until eventually the post ceases to be active (expires).

The situation is entirely different when looked at from the point of view of the users. The pattern of activity of every user is followed over time, a part of the pattern is shown in Fig. 1a. The user indexes are ordered by the time of their first appearance in the dataset, hence the top boundary of the plot indicates the appearance of new users, relative to the beginning of the dataset. The profile of the top boundary shows that new users arrive in "waves", related with the daily cycles. Moreover, the arrivals of new users boost the activity of previous users, which is manifested in the increased density of points in depth of the plot below each "wave". This feature of the dynamics is utilized for designing the model rules in Section 3. Some quantitative measures of the temporal patterns of users and posts in Fig. 1a, b are given in relation to the time-delay and the lifetime distributions, shown in Fig. 3c, d.

Further regularities of the user activity over posts can be extracted from this empirical dataset. Here we focus on some properties of user- and post-dynamics that are relevant for our agent-based modeling. The extracted features are shown in Figs. 2 and 3. In particular, the robust patterns of the blogging dynamics leads to the following conclusions:

- *User interests shift daily towards different posts.* In Fig. 2 the snapshots of the activity within one user community on the weighted bipartite network of discussion-driven Digg stories are shown for three consecutive days. The network is constructed from the subset of the discussion-driven Diggs and keeping only very active users (with more than 100 comments). With the eigenvalue spectral analysis methods [26] three communities on that are identified [13]. For the Fig. 2 the nodes belonging to one of these communities, g_2, which consists of 236 nodes (users and posts) are selected and then the weighted bipartite subnetworks are constructed for each consecutive day. Thus, the weight of the link in Fig. 2 corresponds to the number of comments written by the user on the post *on a given day*, while the color of the link indicates overall emotional contents of these comments (black—negative, white—neutral and red—positive).
- *Universality in the distributions of time-delay of user actions and lifetime of posts.* The delay-time distribution $P(\Delta t)$ is directly related with the user activity pattern, cf. Fig. 1a: for a given user (fixed user index on y-axis) the delay time Δt is defined as the distance between two subsequent points along the time axis. The distribution is then averaged over all users in the dataset. In Fig. 3c we present the delay-time distributions separately for the comments carrying negative/positive emotion. Upper curves are for the popular Digs dataset, while the lower two curves are for the emotional comments on popular Blogs. Apart from the comment polarity, the action delay also strongly depends on daily fluctuations, with the most pronounced peak on the first day (circadian cycles). Note that similar power-law distributions of the time-delay action are found in many other examples of human dynamics [8,27,9]. In the case of posts activity pattern, the quantity that is of interest in this work (see Section 3) is the lifetime of posts, t_p. The lifetime of a post is given by the distance between the first and the last point on the time axis for a given post index, cf. activity pattern in Fig. 1b. Considering all posts in the dataset, the distribution of the lifetime of posts $P(t_p)$ is constructed, and shown in Fig. 3d (data are logarithmically binned). The distribution exhibits a peak at approximately 576 time bins, corresponding to two days of real time, and a power-law tail for longer times. The fitted exponent in that range is 1.12 ± 0.04.
- *Popularity of posts varies at large scale.* The power-law decay of the $P(t_p)$ distribution suggests strong heterogeneity in the lifetime of posts. Further information that can be inferred from the posts' activity patterns is the number of events occurring at a post (post popularity), which is given by the number of points between the appearance of the post and the last event related to it. For each post two different activity patterns can be identified, as mentioned above, i.e., the activity within initial T_0 time bins, and beyond. The parameter T_0 is roughly estimated as the width of the time window, during which new posts were 'exposed' (dense points area in Fig. 1b). It also corresponds to the peak in the lifetime distribution of posts, that is $T_0 \approx 2$ days for this dataset. When T_0 is fixed, then the probability $\mu(T_0)$, that a user looks to a post

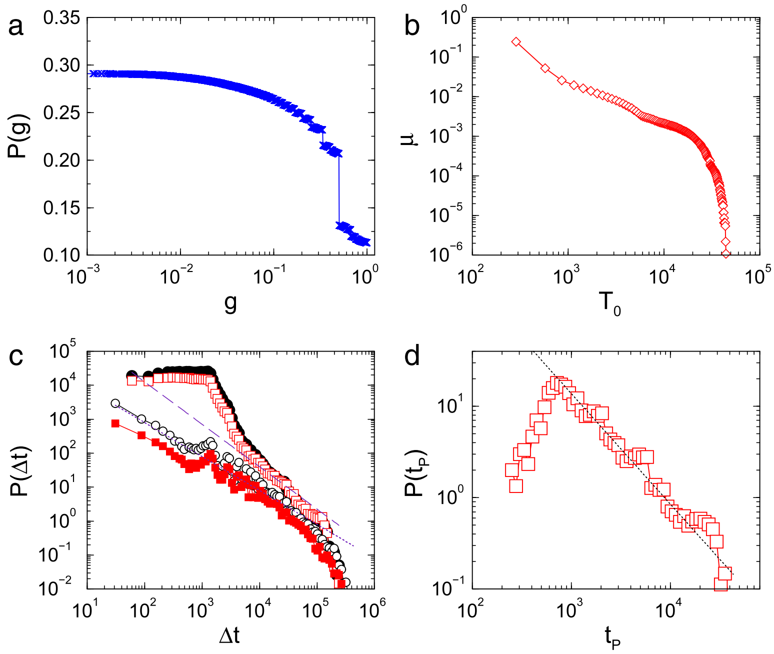

Fig. 3. (a) Distribution of g—the fraction of new posts per user, relative to all posts on which that user was active, averaged over all users in the dataset. (b) Probability μ that a user looks at a post which is older than the specified time window T_0 time bins, averaged over all users and plotted against T_0. (c) Distribution of the time-delay Δt measured in minutes between two consecutive user actions with positive (red) or negative (black) comments, averaged over all users in the dataset. Upper two curves are for popular Diggs, lower curves for popular BBC Blogs. (d) Distribution of the lifetime of posts t_P, averaged over all posts in the dataset, time axis in bins corresponds to $t_b = 5$ min of real time. Data in Fig. (c) and (d) are logarithmically binned with small base 1.1. Straight line indicates slopes, which are explained in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which is older than T_0, can be extracted from the data. Specifically, for each post, this probability is given as the fraction of points beyond the dense area in the posts' activity pattern until the post expires, cf. Fig. 1b. Then for the whole dataset we have $\mu(T_0) = \frac{1}{N_P}\sum_{p=1}^{N_P}\left(\frac{1}{t_p}\sum_{t_{kp}>t_{0p}+T_0}^{t_p} 1\right)$, where N_P is the number of posts, t_p is the expiry time of the post p, while t_{kp} and t_{0p} indicate the moments of the activity at the post p and its creation time, respectively. For the ddDiggs dataset the parameter $\mu(T_0)$ is computed by fixing T_0 in a range of values, and plotted in Fig. 3b.

- *Innovativeness of the users is heterogeneous.* In the data we have full information about every action of each user. Therefore, by looking at the activity list of a given user, we can determine the fraction g of new posts that the user posted, out of all posts on which the user were active in the entire dataset. The values appear to vary over time and users. The distribution $P(g)$ averaged over time and all users in the dataset is shown in Fig. 3a.

As mentioned in the Introduction, these observations about the user behavior over posts are utilized in designing the rules and parameters of the agent-based model in Section 3. In the remaining part of this section we consider some global features of the user dynamics in the empirical dataset: (i) the time series of (emotional) comments, and (ii) the structure of the bipartite network of users and posts that represents connections in the analyzed dataset.

2.2. Prevalence of negative comments on popular posts

Having high temporal resolution data, we can extract fluctuations in the number of comments occurring within a small time bin. Similarly, we can separate the comments according to their emotional contents, e.g., valence polarity. The time series of the number of comments $N_\pm(t)$ with positive/negative emotion valence extracted from the considered dataset of ddDiggs are shown in Fig. 4, where the time bin $t_{bin} = 5$ min is used. Shown also is the time series of the "emotional charge", defined as the difference $Q(t) = N_+(t) - N_-(t)$ at each time bin. As the Fig. 4 shows, the dataset of popular Diggs exhibits an excess of comments with negative emotion valence! Similar features are found in the data of Blogs [12,13]. Whereas, the communication dialogs in the social network MySpace are found to be dominated by positive emotion valence [28]. It is therefore challenging for the theoretical models to unravel the mechanisms that lead to the prevalence of negative emotion in the blogging dynamics. We address this question within the agent-based modeling in Section 3.

Related with the user activity [13], the time series of the number of emotional comments in the discussion-driven popular Diggs data exhibit a fractal structure superimposed to the daily cycles, as shown in Fig. 4. The power-spectrum of $1/\nu^{\phi}$-type, shown in the upper panel in Fig. 4, and the peak corresponding to the daily periodicity. The power spectrum is correlated over the range of frequencies below certain threshold indicated by the arrow, which correspond to times above approximately 2 h

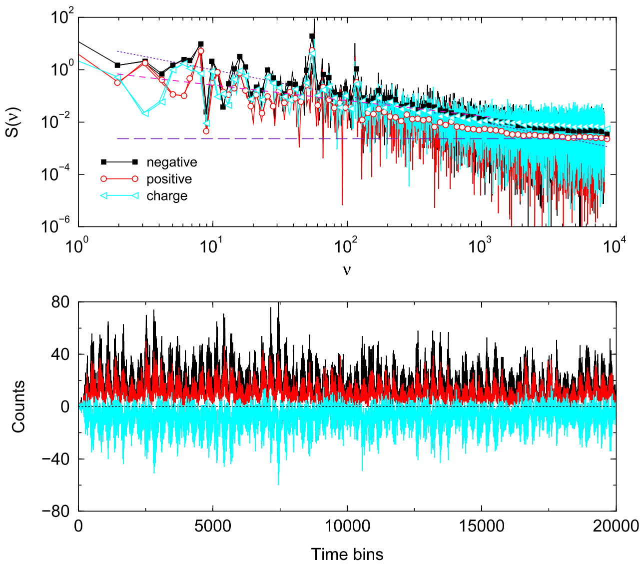

Fig. 4. Time series (lower panel) and their power-spectra (upper panel) of the number of comments $N_{\pm}(t)$ carrying positive emotion (red) and negative emotion (black), and the time-series of "charge", $Q(t) = N_{+}(t) - N_{-}(t)$, of the emotional comments (cyan). Straight lines indicate slopes in the correlation regions, described in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the time domain. The horizontal line in the range of higher frequencies (shorter times) indicates the white noise spectrum. Smooth curves obtained by logarithmic binning are shown with open symbols. The slopes of the correlated parts of the spectrum (indicated by the straight lines) are estimated as $\phi = 0.98 \pm 0.08$ for negative comments and $\phi = 0.86 \pm 0.08$, for the positive comments, in the range [12:2146] of the frequency index. The correlation coefficient is -0.86. The time series of charge fluctuations appear to be very weakly correlated.

2.3. Structure of bipartite networks from the empirical data of Diggs

The high-resolution data in which we have full information about users' IDs and their comments on identified posts, are mapped onto *bipartite networks* with the users, as one partition, and posts (and comments), as the other partition. As discussed in detail in Refs. [9,12,13], upon mapping these networks are then analyzed either as monopartite projections or compressed weighted bipartite networks, in which the weighted links represent the number of comments of a given user to a linked post. For the purpose of this work, the network mapping of the subset of the popular discussion-driven Diggs is considered. For completeness, we also consider the topology measures obtained from the complete dataset (i.e., including the unpopular posts, and the popular posts which remained outside the discussion-driven dataset). We focus on the topology measures as the degree distributions of each partition and the mixing patterns of the weighted bipartite representation of the dataset. These quantitative measures of the topology will be compared later with the ones from the simulation in Section 3.2.

The results of the topology analysis are given in Fig. 5a, b. The degree distributions appear to be specific for nodes in each partition, a feature also found in some other bipartite networks representing the empirical data of techno-social interactions [29]. Specifically, the broad distributions for user-nodes and for post-nodes are dominated by different type of cut-offs. They can be approximated by the following mathematical expressions: the power-law with the exponential cut-off

$$P(q_u) = C_u q_u^{-\tau} e^{-\lambda q_u}; \tag{1}$$

for the user-node degree q_u, and a q-exponential distribution with a power-law tail above the threshold degree σ,

$$P(q_p) = C_p \left[1 + \left(\frac{q_p}{\sigma}\right)\right]^{\theta}; \tag{2}$$

for the post-node degree q_p. Note that the observed differences in the degree distributions reflect the statistical diversity in the role that the nodes of each partition play on the network.

The data are fitted with the mathematical expressions in Eqs. (1)–(2) using Maximum-Likelihood Estimator (MLE) methods. All fits passed the χ^2-goodness test. For the complete dataset (all-data), the best fit parameters are estimated as follows: $\tau = 1.454 \pm 0.005$, $\lambda = 0.00279 \pm 0.00004$, for the user-degree distribution. While $\theta = 6.5 \pm 0.1$ and $\sigma = 480 \pm 20$, for the the distribution of posts degree (fitted range $\in$ [10, 1000]). Note that in the case of popular posts, the distribution of post-degree does not contain the posts of low degree (strength), by definition. Whereas, for the users attached to these popular posts the distribution of user-degree exhibits the same exponent, $\tau = 1.454 \pm 0.05$, and a shorter cut-off $1/\lambda$, with $\lambda = 0.011 \pm 0.002$.

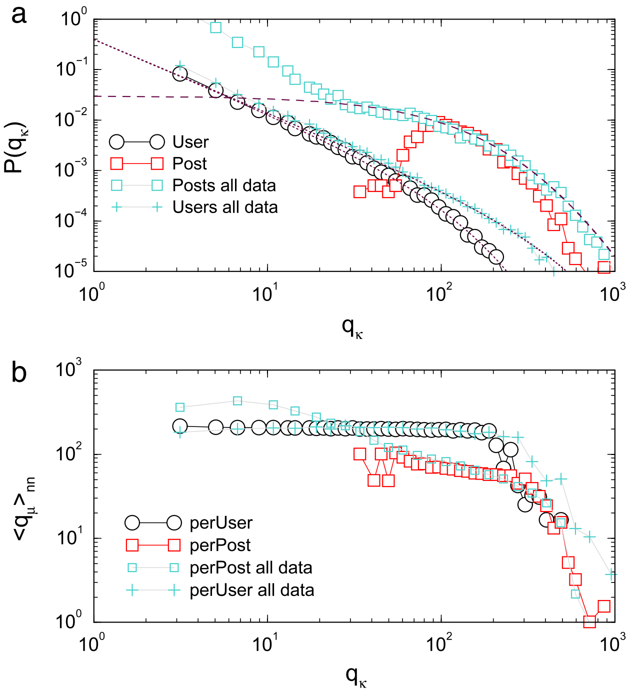

Fig. 5. The degree distribution of user-nodes (◯) and post-nodes (□) of the network of popular discussion-driven Diggs, panel (a), and the network assortativity measures, panel (b). Shown are also the respective quantities computed from the whole dataset, including the posts with normal popularity, indicated by "all data" in the Legend. Fits according to expressions (1) and (2) and explained in the text, are shown by dotted and dashed lines, respectively.

The (dis)assortativity is a topology measure related with the link correlations in a node's near neighborhood on the network. When averaged over all nodes, it may result in a systematic trend in the node linking pattern. In particular, the trend can be either assortative, where linking occurs between the nodes of similar degree, or disassortative, linking with dis-alike nodes, or none. On the bipartite network, the node neighbors are by definition the nodes of the other partition. Therefore, two types of assortativity measures are defined, depending on whether they are looked at from the user-nodes ("perUser") or from the post-nodes ("perPost"). In particular, we have computed the network-average degree of the post nodes which are linked to the user of a given degree, and vice versa, the average degree of the user nodes linked to the post of a given degree. The results are shown in Fig. 5, top panel. In analogy to the degree distributions, we have also computed the assortativity measures for the complete set of data (including unpopular and non-discussion Diggs). These results suggest the mixing pattern with slight dis-assortativity of the post-nodes: the popular posts contain links to users who are on average not the most active users. The same trend is found in less popular posts until the degree drops below ∼25, corresponding to the inflection point of the post-degree distribution. On the other hand, the user-linking pattern shows virtually no assortativity (the slope of the curve is close to zero), which suggests that *on average* users link equally to all kinds of posts.

In the following section we will introduce an agent-based model of blogging, where the agents' actions on posts will give rise to the bipartite network of a similar structure.

3. Agent-based model of emotional blogging

Following Ref. [24], we assume that the individual emotional state (arousal and valence) of each agent can be described by two nonlinear equations, which are the subjects of the environmental fields. For our system on bipartite networks, the arousal and the valence are associated with each user-node(!) and their values, kept in the intervals $a_i(t) \in [0, 1]$ and $v_i(t) \in [-1, 1]$, are updated according to the following nonlinear maps (case (a)-applies if $\Delta t_i < 1$, case (b)-otherwise):

$$a_i(t+1) = \begin{cases} (1-\gamma_a)a_i(t) + [h_i^a(t) + qh_{mf}^a(t)](d_1 + d_2(a_i(t) - a_i(t)^2))(1 - a_i(t)) & \text{(a)} \\ (1-\gamma_a)a_i(t) & \text{(b)} \end{cases} \quad (3)$$

and

$$v_i(t+1) = \begin{cases} (1-\gamma_v)v_i(t) + [h_i^v(t) + qh_{mf}^v(t)](t)(c_1 + c_2(v_i(t) - v_i(t)^3))(1 - |v_i|) & \text{(a)} \\ (1-\gamma_v)v_i(t) & \text{(b)} \end{cases} \quad (4)$$

where $i = 1, 2 \cdots N_U(t)$ indicates the index of user node and t-the time bin. The coefficients d_1, d_2 and c_1, c_2 characterize the maps themselves, while the network environment effects appear through two types of fields: the local fields $h_i^a(t)$ and

$h_i^v(t)$, and the mean fields $h_{mf}^a(t)$ and $h_{mf}^v(t)$. Note that the local fields $h_i^a(t)$ and $h_i^v(t)$ vary not only in time but also from user to user, depending on their connections on the network, and due to the evolution of the network itself (see details below). Whereas, the mean fields $h_{mf}^a(t)$ and $h_{mf}^v(t)$ may act on a larger number of users, while also fluctuating in time. In our model they steam from currently active posts and, thus, may be seen by all users who are attached to these posts. The mean fields indicate how the overall activity moves through posts, as a kind of "atmosphere" at the Blog site. The contribution from the mean fields in our model is taken with a fraction $0 \le q \le 1$, which is varied as a free parameter in Eqs. (3) and (4), and it is added to the contributions from the local fields.

The fields $h_i^a(t)$ and $h_{mf}^a(t)$ in Eq. (3), which affect user i's arousal at step $t+1$, are determined from the posts in the *currently active part of the network*, $\mathscr{C}(t, t-1)$, along the links of that user. Specifically,

$$h_i^a(t) = \frac{\sum_{p \in \mathscr{C}(t,t-1)} A_{ip} a_p^{\mathscr{C}}(t)(1 + v_i(t) v_p^{\mathscr{C}}(t))}{\sum_{p \in \mathscr{C}(t,t-1)} A_{ip} n_p^{\mathscr{C}}(t)(1 + v_i(t) v_p^{\mathscr{C}}(t))}; \qquad h_{mf}^a(t) = \frac{\sum_{p \in \mathscr{C}(t,t-1)} a_p^{\mathscr{C}}(t)}{\sum_{p \in \mathscr{C}(t,t-1)} n_p^{\mathscr{C}}(t)}, \tag{5}$$

where $a_p^{\mathscr{C}}(t)$ and $v_p^{\mathscr{C}}(t)$ are the total arousal and the average valence of the post p calculated from the comments in two preceding time steps, while $n_p^{\mathscr{C}}(t)$ is the number of all comments posted on it during that time period. A_{ip} represents the matrix elements of the network, i.e., $A_{ip} > 0$ if user i is connected with the active post p, while $A_{ip} = 0$ if there is no link between them at the time when the fields are computed. Note that such links may appear later as the system evolves. In Eq. (5) the individual arousal fields $h_i^a(t)$ is modified by (dis)similarity in user's actual valence, $v_i(t)$, and the valence of recent comments on the post, $v_p^{\mathscr{C}}(t)$.

Regarding the valence fields in Eq. (4), we take into account contributions from the positive and the negative comments separately, while the neutral comments do not contribute to the valence field. Depending on the current emotional state of the agent, positive and negative fields can lead to different effects [24], in particular, a positive (negative) state will be influenced more with a negative (positive) field, and vice versa. Here we assume that both components influence user valence, but with different strengths according to the following expression:

$$h_i^v(t) = \frac{1 - 0.4 r_i(t)}{1.4} \frac{\sum_{p \in \mathscr{C}(t,t-1)} A_{ip} N_p^+(t)}{\sum_{p \in \mathscr{C}(t,t-1)} A_{ip} N_p^{emo}(t)} - \frac{1 + 0.4 r_i(t)}{1.4} \frac{\sum_{p \in \mathscr{C}(t,t-1)} A_{ip} N_p^-(t)}{\sum_{p \in \mathscr{C}(t,t-1)} A_{ip} N_p^{emo}(t)}, \tag{6}$$

where the valence polarity of the user i is given by $r_i(t) = \frac{v_i(t)}{|v_i(t)|}$, and $N_p^{\pm}(t)$ is the number of positive/negative comments written on post p in the period $[t-1, t]$. The normalization factor $N_p^{emo}(t)$ is defined as $N_p^{emo}(t) = N_p^+(t) + N_p^-(t)$. The mean-field contributions to the valence steam from the entire set of currently active posts $\mathscr{C}(t, t-1)$, and are independent of how users are linked to them:

$$h_{i,mf}^v(t) = \frac{1 - 0.4 r_i(t)}{1.4} \frac{\sum_{p \in \mathscr{C}(t,t-1)} N_p^+(t)}{\sum_{p \in \mathscr{C}(t,t-1)} N_p^{emo}(t)} - \frac{1 + 0.4 r_i(t)}{1.4} \frac{\sum_{p \in \mathscr{C}(t,t-1)} N_p^-(t)}{\sum_{p \in \mathscr{C}(t,t-1)} N_p^{emo}(t)}. \tag{7}$$

However, the mean-field effects are perceived individually by each user, depending on the polarity $r_i(t)$ of the user's current valence. Note that in the Eqs. (6)–(7) the prefactor 0.4 multiplying the indicator of the positive/negative valence $r_i(t)$, is an arbitrary value chosen to interpolate between the extreme cases 0 and 1, and away from the fully symmetrical situation with the prefactor 0.5.

3.1. Dynamic rules and control parameters

Theoretical modeling of the complex behavior on the Blogs faces very tough requirements. In this section, the rules of agents interactions on the network are formulated in view of user behavior on real Blogs and Diggs and the observations from the quantitative analysis of the related empirical data. In particular, the dynamic rules and parameters of our model are motivated by the temporal patterns in Fig. 1a, b and the properties summarized in Figs. 3. Moreover, additional features of the ddDiggs data, shown in Figs. 2 and 4, suggest the dynamics with dominance of the negative emotions and with user's focus systematically shifting towards different posts. In the implementation of the model rules, we also make use of some general features of human dynamics, i.e., the occurrence of circadian cycles and delayed action to the events, mentioned in the Introduction, and assume that the arousal drives an action, as commonly accepted in the psychological literature.

The rules are implemented in the C++ code as follows. The system is initialized with typically 10 Users who are connected to 10 Posts, to start the lists of the *exposed* and the *active* posts and the *prompted* and the *active* users. Then at each time step:

- The system is driven by adding $p(t)$ new users (note the correspondence of one simulation step with one $t_{bin} = 5$ min of real time); Their arousal and valence are given as uniform random values from $a_i \in [0, 1]$, $v_i \in [-1, +1]$, then updated with the actual mean-field terms. By the first appearance each user is given a probability $g \in P(g)$ to start a new post. The new users are then moved to the *active user list*;

Table 1
Control parameters of the agent-based model of emotional blogging.

Local maps	Agents & posts properties	Driving
$c_1 = d_1 = 1$[a]	$t_P \in P(t_P)$[b]	$\langle p(t)\rangle = const = 6$[b]
$c_2 = 2.0$[a]	$\Delta t \in P(\Delta t)$[b]	$q = 0.4$[a]
$d_2 = 0.5$[a]	$T_0 = 576$ tbins[b]	$a_0 = 0.5$[a]
$\gamma = 0.05$[a]	$\mu(T_0) = 0.05$[b] $g \in P(g)$[b]	

[a] Values within theoretical limits.
[b] Values inferred from the empirical data, cf. Figs. 1 and 3.

- The emotional states for all present users are relaxed with the rate γ, according to the second row in the Eqs. (3) and (4);
- The network area $\mathscr{C}(t, t-1)$ of the *active posts* is identified as a post on which an activity occurred in two preceding time steps; then the lists of *active users* is updated from the users linked to these posts, as follows:
  – Users linked to the active posts are considered as *exposed* to the posted material and decide when they will act on it, i.e., they are given new delay-time from the distribution $P(\Delta t)$; All users whose current delay time $\Delta t < 1t_{bin}$ are *prompted* for update the emotional states according to the first rows in Eqs. (3) and (4), with their actual network fields computed from the Eqs. (5)–(7). An updated user is moved to the *active user list* with the probability $a_0 a_i(t)$ proportional to its current arousal, else it gets a new delay time $\Delta t \in P(\Delta t)$;
- Every active user:
  – adds a new post with the probability g or otherwise comment to one of the *exposed posts*, which are not older than T_0 steps; Users are linked to posts preferentially with the probability $p_p(t) = \frac{0.5(1+v_p^{\mathscr{C}}(t)v_i(t))+N_p^c(t)}{\sum_p [0.5(1+v_p^{\mathscr{C}}(t)v_i(t))+N_p^c(t)]}$, depending on the number of comments on it $N_p^c(t)$ and the valence similarity;
  – and with probability μ comments a post which is older than T_0 steps. The post is selected preferentially according to the negativity of the charge of all comments on it, with (properly normalized) probabilities $p_{j,old}(t) \sim 0.5 + |Q_j(t)|$, if the charge is negative, else $p_{j,old}(t) \sim 0.5$; Lifetimes of the posts are systematically monitored (already expired posts are not considered);
  – Current values of the valence and arousal of the user are transferred to the posted comment or the new post; User is given a new delay-time $\Delta t \in P(\Delta t)$; New posts are given lifetime $t_P \in P(t_P)$.
- Delay-time Δt for all other users is decreased by one. Time-step closes with updating the lists of the exposed and the active posts, and the lists of the exposed and the prompted users.

According to the above dynamic rules of the model, one can identify the parameters which control the dynamics at different levels. In particular, we use the following parameters, distributions or time-series which characterize the local maps, the agent's and post's properties, and the driving conditions. They are summarized in the Table 1.

A comment regarding the parameters is in order at this point (see also more discussion in Section 4, where the simulations for varied parameters are carried out). As stated in the Introduction, most of the parameters of the agent's dynamics can be inferred from high-resolution data, such as our dataset of Diggs. The numerical values of the parameters of the model, which are inferred from the empirical data of ddDiggs, are listed in the Table 1 and the distributions $P(g)$, $P(\Delta t)$ and $P(t_P)$ are shown in Fig. 3. Strictly speaking, the values of the control parameters will depend on the empirical dataset considered. Specifically, the parameters as the lifetime of posts, t_P, and users inclination to posting new posts, g, as well as the probability of choosing an old post, $\mu(T_0)$, strongly depend on the dataset. Note also that they might have hidden inter-dependences in view of the nonlinear process underlying the original dataset. For instance, if on a certain Blog site users are more inclined towards posting new material, which would yield increased probabilities of large g, then the lifetime of posts may decline, resulting in a steeper distribution. Therefore, *it is important to derive these parameters from the same dataset in order to ensure their mutual consistency*. Although our model works for a wide range of parameter values, here we keep the parameters extracted from the empirical data of the popular discussion-driven Diggs in order to enable a comparison of the results to the largest possible extent.

In contrast to the above empirical values of the parameters, the relaxation rate of the arousal and the valence γ and the parameters d_1, d_2, c_1, c_2 of the maps in Eqs. (3)–(4) can not be extracted from this type of empirical data. In the simulations these parameters are kept within theoretical limits [30]. Specifically, the nonlinear maps in Eqs. (3)–(4) have fixed points, respectively in the area $a \in [0, 1]$ and two fixed points $v \in [-1, 0]$ and $v \in [0, 1]$. The position of these fixed points depends on the values of the parameters and the coupled fields $[h_i^a(t) + qh_{mf}^a(t)]$ and $[h_i^v(t) + qh_{mf}^v(t)]$. The theoretically reasonable parameters are then those for which the fixed points of the maps do not fall at the border of the available phase space for typical values of the fields. By examining the fluctuations of the fields in our simulations, we set the acceptable values of these parameters, which are listed in Table 1. Note that choosing nonzero values c_1, d_1 only contributes to the field calibration, therefore we can set them to unity. The case with zero values for c_1, d_1, on the other hand, would lead to a different situation, where the fields can affect an agent state only via the nonlinear terms, which we think is unrealistic.

3.2. Simulation results: the case with constant flux of agents

In this paper we restrict the discussion to the case where the flux of agents is constant in time. Note that for the comparison with the empirical data, the time step of the simulations corresponds to one time bin of the real data. We keep

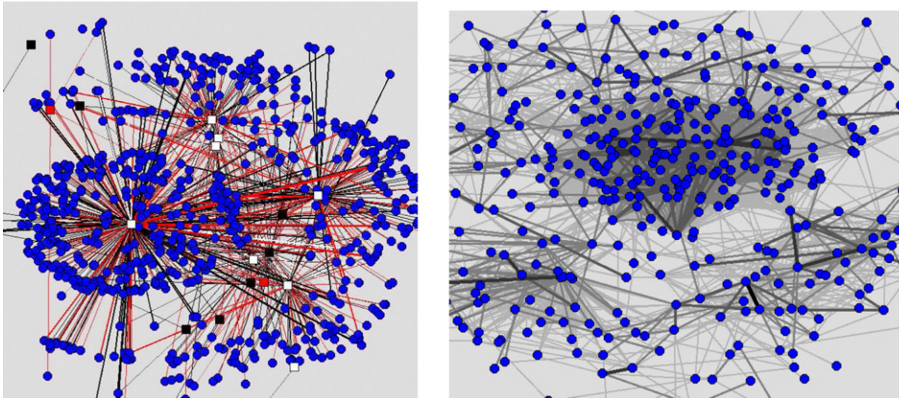

Fig. 6. Left: an example of the emergent bipartite network with agents (•) and posts (□) after the first 128 simulation steps. The weighted links represent the number of comments, while their color indicates overall emotion valence: red (positive), black (negative), and white (neutral). Right: a part of the larger network obtained after 4032 time steps in the simulations when the parameter $q = 0$ (see Section 4), and projected onto agent partition. For better vision shown are only agents with the strengths larger than 3 which belong to three smaller communities identified on this network. The weights of the links, indicated by gray scale, represent the common number of posts per pair of the agents. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

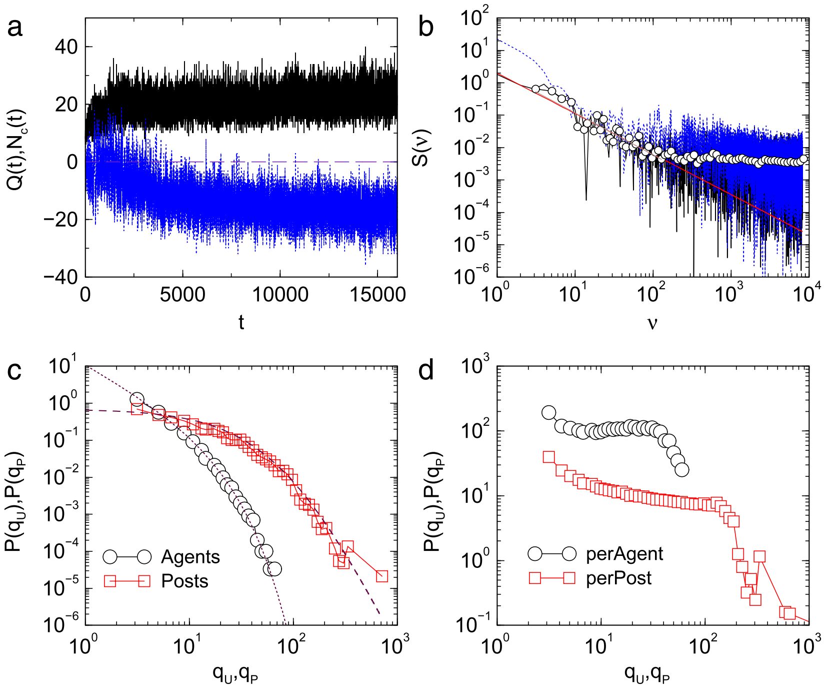

Fig. 7. Simulation results obtained for the system of the emotional agents is driven by adding a constant number of agents $\langle p(t)\rangle = 6$ per time step: time-series of the number of comments and charge (a) and their power-spectra (b). Degree distributions of agent-nodes and post-nodes, $P(q_u)$ and $P(q_p)$ with the fit lines according to the expression (1) and (2), are shown in panel (c) and the related assortativity measures in panel (d).

the time units in bins of 5 min of the driving signal $p(t)$, which gives the value of the average $\langle p(t)\rangle$ users per time bin, given in the Table 1. Other possibilities can be easily implemented into model rules, see Ref. [30]. In order to compare the dynamics of the emotional agents with the results observed in the empirical data in Section 2, in the simulations we sample the same quantities related with (i) fluctuations of the agent's activity in time, and (ii) the network structure that emerges through the actions of the agents. The simulation results are then analyzed in full analogy with the empirical data. For the illustration, the bipartite network that emerges in the initial 128 steps is shown in Fig. 6. The simulations are extended till 16,384 time steps, corresponding to roughly 57 days of real time. The results of the analysis of the simulated data are summarized in Fig. 7(a–d).

The simulated time series of the number of comments made by the agents at every time step, and the emotional charge of these comments are given in Fig. 7(a). As the figure shows, in the process of the agents linking to the posts, the excess negative charge sets in after some initial fluctuations, and remains negative for a very long time. In this way the model dynamics reflects the global feature of the user-activity on posts, observed in the analysis of the empirical data in Section 2. Furthermore, despite the absence of the circadian cycles, which are present in the empirical time series, the simulated time series also exhibit a fractal structure. The power spectra of these time series are shown in Fig. 7(b). In the range of frequencies indicated by the straight line in Fig. 7(b), the temporal correlations of the type $1/\nu^{\phi}$ can be seen, similar to the ones observed in the empirical data in Section 2. The exponent $\phi = 1.29 \pm 0.10$ is the best fit of the slope in the range [2:128] of the frequency index. Whereas, white noise signal is found for $\nu > 128$ large frequencies (short times), corresponding to $\phi = 0$. Hence, the system's internal dynamics is capable of building a certain cooperative behavior over large temporal scales, despite the absence of correlations in the external inputs.

Next we analyze the network of agents and posts that emerges after long simulation times. The initial stage of the network evolution is shown Fig. 6. The multiplicity (weight) of the links indicate the number of comments of the agent to the post, and the color of the link, as explained in the caption to Fig. 6, indicates the cumulative emotion valence of all comments along the link in the preceding period. According to the model rules, the network evolves with the addition of agents and the addition of posts, and with the addition of new links, or increase of the widths of old links between the agents and previously existing posts. As the Fig. 6 shows, at the network level the amount of positive (red) links is well balanced with the negative (black) ones in the initial stage of the network growth, which is in agreement with the initial part of the time series of charge in Fig. 7(a). The dominance of negative (black) links sets-in in the later stages of the network evolution. The structure of the network grown within 16,384 time steps is analyzed. The degree distributions of the agent- and the post-partition are shown in Fig. 7(c) and the corresponding assortativity measures in Fig. 7(d). The distributions for the agent-degree and post-degree are fitted with the same expressions as the ones in the case of empirical data, Eqs. (1) and (2), respectively. Apart from the cut-off values, the distributions are reasonably close to the ones obtained from the empirical data. The fits with the MLE method give the parameters $\gamma = 1.546 \pm 0.017$ and $\lambda = 0.11 \pm 0.07$ for the agent-degree distribution, and $\theta = 5.78 \pm 0.06$ and $\sigma = 84 \pm 2$, for the post-degree distribution. Moreover, the assortativity results exhibit the same tendencies (for the respective degree larger than 10) as the corresponding measures in the empirical data.

4. Effects of varied parameter values

In the driven nonlinear dynamics of the interacting agents in our model, the control parameters are an essential part of the story. Varying the control parameters may influence the outcome global states in different ways. In this section we explore how the topology of the network is changed by varying one of the key driving parameters—the fraction $q \in [0, 1]$ of the mean-field contribution to the agent's emotion state. But before presenting the simulation details, we would like to briefly comment on all other model parameters, which are listed in the Table 1.

The parameters marked by (^{b}) in the Table 1 have their realistic values for that particular dataset and the time window where the data are collected. As mentioned above, the set of these parameters (i) have the values determined in an experimental system, therefore they can not be considered as free parameters; (ii) their values are mutually consistent, in view of possible interdependences, as in the case $\mu(T_0)$, which is clear by definition, but other hidden dependences may also occur. For instance, for a given set of users, the user's inclination to posting new posts, g, may affect the distribution of the lifetime of posts by keeping the same level of activity. Similarly, if the delay-time distributions are more flat, the average level of the agent's activity would increase, which may lead to increased connectivity of the network by fixed distribution of the lifetime of posts. Therefore, these parameters should not be varied independently. When new empirical data is considered, the whole set of the parameters need to be computed in order to ensure their consistent values. In Section 2 we have shown how these parameters can be computed from any high-resolution dataset.

The remaining parameters, marked by (^{a}) in the Table 1, can not be estimated from the available empirical data gathered at Blog sites. For determining their "natural" values in a real user system, one would need additional information, which can be obtained by targeted social psychology measurements. On one side, these parameters concern the psychology profile of the Blog users (i.e., the local map parameters c_2, d_2, γ), and on the other, the estimation of the overall "atmosphare" on the Blog (q, a_0), which may affect the users behavior. To our knowledge, results of such measurements with Blog users are not currently available. Therefore, these parameters appear as free parameters in the model. In the simulations their values are kept within theoretical limits, given the mathematical formulation of the model. Specifically,

- $c_2 > 0$ and $d_2 > 0$ are the nonlinearity parameters of the maps, which characterize each individual agent. We keep their values such that for the typical values of the fields the fixed point of the map does not fall at the corners of the allowed area of the phase space [30]. For simplicity, we assume equal parameters for all agents (maps). Further distinction between the agents is possible when (c_2, d_2) are varied form agent to agent. For such purposes, however, one would also need a more clear guide from psychology. Note that theoretically the upper limit of the parameters is not fixed, however, as soon as c_2 (or d_2) is too large, the map reaches quickly its fixed point, where it tends to stay (apart from the relaxation). To avoid such unrealistic situations for typical field fluctuations in our model, these nonlinearity parameters are kept in the range [0.5, 4].
- γ is another parameter for which no empirical data are currently available. It is also related to the nonlinear maps in that it affects each step by retracting the mapped point back towards zero. When this relaxation rate is too big, the map would reach zero practically after each time step, which we think is not realistic for the emotional discussions that our model aims to describe. Otherwise, if the rate is too small, the relaxation would take long time, which appears unrealistic

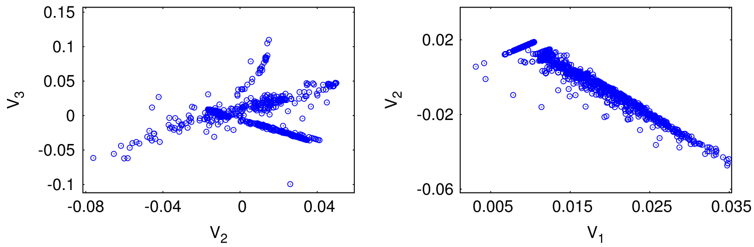

Fig. 8. Community structure detection in the case when the mean-field is absent, $q = 0$ (left), and when the whole interaction is due to mean-field, $q = 1$ (right).

when compared with the temporal patterns of the user activity in Fig. 1. In this situation the agents' arousal would keep large values for a long time, making them more active on average than the users of the comparable dataset. Namely, the computed average time between user actions in the empirical data is approximately 674 min, which includes also average waiting times. Thus, in the simulations we take a plausible value $\gamma = 0.05$, which corresponds to characteristic relaxation time $1/\gamma = 20$ time bins, i.e., 100 min in real time. Again, a targeted measurement of the emotion relaxation would be necessary to estimate this parameter for a given user community. Possible differences between users in this respect can be also implemented in our model.

- a_0 is a parameter by which one can tune "responsivity" of the environment. Theoretically, it scales the arousal effects at the individual agent level. In the collective states this will be manifested in a fuzzy lower bound of the arousals by which the actions took part. We think that this is more natural than introducing a sharp threshold when the arousal induces the action. Cf. examples of the circumplex maps in our work in Ref. [30].

4.1. Tuning the community structure

The above discussed parameters control the *individual emotion dynamics* of each agent. Thus their effects on the global states may be primarily manifested in the phase space of the emotion variables. Different effects are expected when one of the parameters which modifies the driving modes of the system is varied, cf. Table 1. Here we simulate the system's behavior for different values of the parameter q—which measures the fraction of mean-field contribution in boosting the agent's emotional state. By definition, the mean-field varies with time, depending on what part of the network was recently active, and it acts equally to all currently active agents. In this case the effects are manifested on the measurable quantities that characterize the collective states, as we demonstrate in the remaining part of this section.

In the case studied in Section 3 we considered the parameter $q = 0.4$, which means that 40% of the influence on an agent's emotional state is due to the "mean-field" part (i.e., from all recently active posts), while the remaining 60% is due to posts directly connected with the agent. According to the model rules, both types of fields steam from the recently active posts, i.e., in time window of two time steps (ten minutes of real time). The motivation to consider such balanced contribution from individual fields and the mean field, is to closely match the situation in the empirical data of the discussion-driven Diggs, where more than 50% of the user actions are comment-to-comment.

In the model the balance between individual and global fields can be modified by changing the parameter q. In order to understand the nature of these two contributions, here we first consider two limiting cases: $q = 0$, corresponding to the situation where the dynamics is driven by the local contacts alone, and $q = 1$, where the mean-field dominates. Sporadically, such extreme situations may occur on real Blogs. An example corresponding to $q = 1$ is the case with externally-driven dynamics, where many users comment the same post (or few posts), with very little discussion among each other. Depending on the emotion expressed in the original post, the individual comments can get negative as well as positive emotion. This kind of events at a given post are well represented by a star-like network structure, as found in previous work [12]. The other extreme, $q = 0$, may be found in the situation when a group of users gets engaged in mutual discussion, with even shifting the subject away from the original post. Such group discussions can often get critical (negative emotion dominates), while the entire community may still have balanced emotions. More realistic situations are thus in between these limiting cases. Therefore, a finite value $q \in (0, 1)$ between these two limits is more realistic to describe the Blog dynamics by the model.

Our simulations in this section show how the mean-field affects the mechanisms leading to the collective behaviors and the occurrence of communities on the network. The results are summarized in Figs. 8–10. In the limit $q = 0$, agent's communications driven by their local fields leading to the diversity among agents, which results in the community structure shown in Fig. 8 (left). Whereas in the limit of strong mean-field, Fig. 8 (right), such diversity is lost, consequently we find that all agents tend to belong to the same community.

Further interesting results are found in correlations between the mean-field fraction q and the range of fluctuations of the activity (the number of comments) and the emotion polarity of the comments. In Fig. 9 the large mean-field contribution $q = 1$ reduces the range of fluctuations in the number of comments (note that the lower bound is 6, in view of adding 6 new agents every time step). While, in the $q = 0$ limit, the range of fluctuation is larger. At the same time, the presence of mean-field facilitates appearance of the excess negative charge. In Fig. 9 (right) a stationary time series of negative charge settles after initial fluctuations. In the absence of the mean-field $q = 0$, however, the charge of the comments at the level of whole

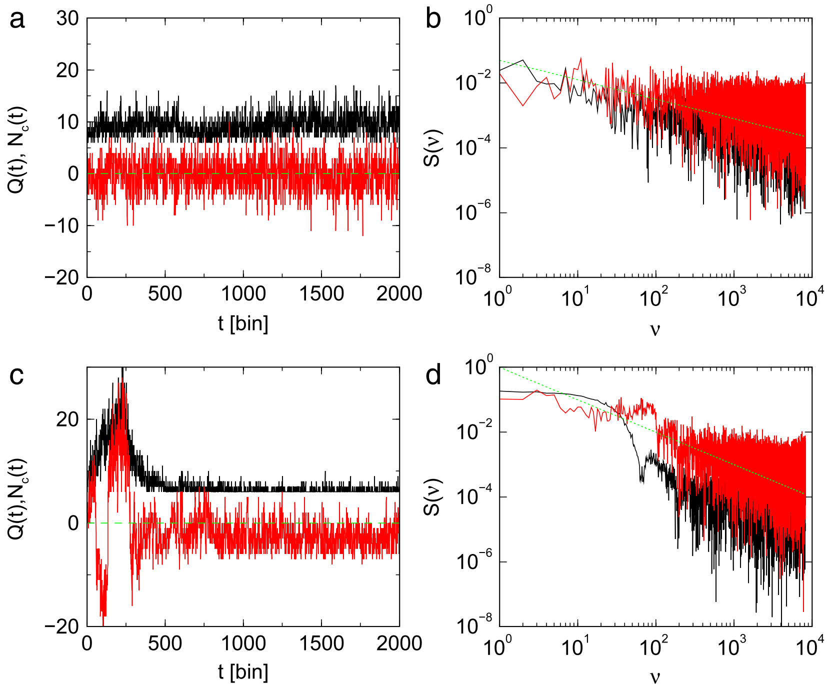

Fig. 9. (a) Time-series of the number of comments $N_c(t)$ (upper/black line) and their charge $Q(t)$ (lower/red line) in the case when the mean-field is absent, $q = 0$, and (b) their power spectrum; (c) and (d) the corresponding time series and their power spectrum when the whole interaction is due to mean-field, $q = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

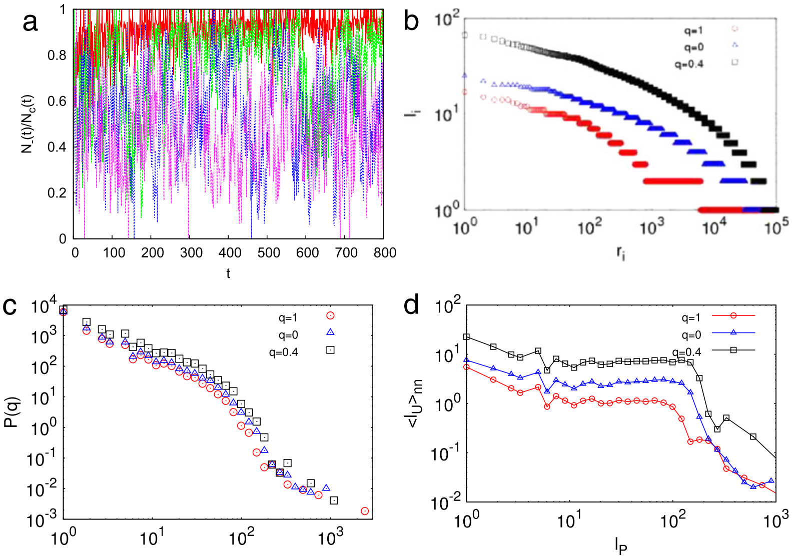

Fig. 10. Comparison of varied fraction q of the mean-field on the time series and the emergent network properties. Panel (a) shows an example of initial time series with the fluctuations in the fraction of negative comments for four different values of $q = 0.8$ (red line), 0.6 (green), 0.4 (blue) and 0.2 (pink). (b) Agent-strength ranking distribution, (c) post degree distribution, and (d) mixing patterns per post, for three values of the parameter q as indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

system remains well balanced with fluctuations around zero, cf. Figs. 9 (left). Note that these time series also have different power spectra, as shown in the top panels in Figs. 9. After having balanced fluctuations in the beginning of the simulations, the charge polarity breaking in our model occurs dynamically when the network grows large enough. The strong mean-field part when $q \lesssim 1$ contributes to homogenization of the agent's communities, thus reducing the diversity of their emotion fluctuations. In conjunction with the local fields, e.g., when $q = 0.4, 0.6$, the agent's diversity is larger leading to local communities which are often linked to some post-node hubs. Due to the agent's preference towards negative comments, such hubs can have a large negative charge, promoting the community growth [30]. In over 90% of the simulation runs performed with different series of random numbers we obtain a negative share at the level of the whole network. This mechanism, however, does not exclude the situation with the majority of positive comments to occur as a collective effect, which we find in less than 10% of the runs. As shown in Fig. 10(a), a typical fraction of negative comments (compared to the number of all comments) at the level of the whole system increases when $q \to 1$. However, the fluctuations are stronger in the situations where true competition between the mean-field and the local fields is behind the network growth mechanisms, for instance when $q = 0.4$ and 0.6. In these cases the local fluctuations of the emotion charge are linked with the occurrence of the communities.

The activity of agents is affected by increased mean-field contribution to agent's emotion dynamics. The above results of the community structure and the fluctuations in the number of comments are in agreement with reduced diversity among agents at strong mean-fields, which is also observed in the emergent network structure. In 10(b) we show how the agents ranking distribution is changed with increased q values. In contrast to the agent's linking pattern, the posts-partition of the network has a similar structure for different q-values. As shown in 10(c, d), the strength distribution of posts, as well as the mixing patterns plotted "per-post", exhibit qualitatively similar functional dependences when the fraction q of the mean-field is varied.

5. Conclusions

In this work we have investigated the blogging dynamics by using two approaches: (i) analyzing the high-resolution empirical data from popular Diggs, and (ii) introducing an agent-based model with the rules and parameters closely related with the same empirical data. Our focus was on the emergence of the emotion-driven collective behaviors on Diggs, Blogs and similar Web portals where the user interaction is mediated by the posts. For this purpose, we selected a set of the empirical data related with *discussion-driven popular posts*, which might be the best candidate where the emotion plays a role in the collective behavior. In the agent-based model, among other properties, the agents have emotion (arousal and valence), which play a key role in their actions. Both the empirical and the simulated data are analyzed in parallel, i.e., using the same quantitative measures defined within the statistical physics of complex systems. Thus the quantitative comparisons between the model and the real system are made possible, which facilitates the estimation of the role of the emotion in the the nonlinear dynamics of the real system. A brief summary of the results indicating the occurrence of the collective behaviors is given below. Furthermore, the model provides insight into the mechanisms underlying the collective behaviors and the ways to influence them.

Our main findings, when the parameters are kept at their "native" values extracted from the empirical data, provide a qualitative agreement between the simulated agents system and the real bloggers. In particular, in the

- *time series* of the emotional comments we find the appearance of the long-range temporal correlations with $1/\nu^{\phi}$ power spectrum, antipersistence and the occurrence of the excess negative charge, similar to the real data;
- *topology of the emergent bipartite network* in several quantitative measures, the exponents of the related degree distributions, mixing patterns, and the mesoscopic structure with the communities, are found in qualitative agreement with the ones obtained in the real system. The observed quantitative differences can provide a measure of the role of other user properties, which are not taken into account by the emotional agents.
- *mechanisms of prevalence of negative comments* are revealed through the model dynamics, which can not be inferred from the empirical data analysis.

We further demonstrate how the community structure can be tuned via varying the balance between the global influence (e.g., due to an external event) affecting all agents, and the local events individual to each agent on the network. This balance in our model is controlled by the fraction $q \in [0, 1]$ of the mean field contribution to the agent's emotional state. While local fields induce diversity among the agents, leading to the occurrence of different communities, increased mean-field fraction tends to homogenize the agents. Consequently, they eventually tend to belong to one community. This also facilitates keeping the negative charge of the emotional comments, once its polarity is dynamically broken. Although the parameter q can not be estimated directly from the information contained in the considered empirical data, the comparison of the system's behavior, as explained above, suggests that the right balance of the local and mean fields in the real system is such that it allows strong fluctuations, dominance of the negative charge and the occurrence of communities. For instance, such situations are found in our model for $q = 0.4$ and 0.6, studied above.

In conclusion, parallel analysis of the high-resolution empirical data and the agent-based modeling with the parameters closely related to that data, demonstrates that the emergent collective behavior of the users acting at popular posts is to a large extent driven by their emotional behaviors. Through the model we gain an insight into underlying mechanisms (i) how such emotional actions of individual agents lead to an emergent collective state with the communities, and (ii) the role of the excess negative emotion in it. We further provide the ways to affect the community formation by tuning the ratio of the local and the global influence on the emotion of individual agents. From the point of view of network theory, this is a unique model where a stable and controllable community structure can be obtained with the (bipartite) network growth rules which are not related with the topology features of the linking nodes, but with their emotion dynamics. We also provide a systematic methodology to extract the consistent set of parameters of the model from a given set of high-resolution data. In this way the model can be used for predictions of potential collective emotional behaviors of users on the Blog site, from which the data are collected.

Acknowledgments

The authors are thankful for the support from the national program P1-0044, the project P1-0044-3 of the Republic of Slovenia and the European action COST-MP0801.

Author's contribution

Designed the research: BT. Performed the simulations: MM. Analyzed the data: BT, MM. Produced the figures: BT, MM. Wrote the paper: BT.

References

[1] M. Francl, Blogging on the sidelines: bloggers shouldn't be relegated to the sidelines of the scientific literature, Nature chemistry 3 (2011) 183–184.
[2] J. Kleinberg, The Convergence of Social and technological Networks, Communications of the ACM 51 (2008) 66–72.
[3] J. Giles, Social science lines up its biggest challenges, Nature 470 (2011) 18–19.
[4] A. Cho, Ourselves and our interactions: the ultimate physics problem? Science 325 (2009).
[5] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Review of Modern Physics 81 (2) (2009) 591–646.
[6] R.D. Malmgren, D.B. Stouffer, A.S.L.O. Campanharo, L.A. Amaral, On universality in human correspondence activity, Science 325 (5948) (2009) 1696–1700.
[7] P.S. Dodds, C.M. Danforth, Measuring the happiness of large-scale written expression: songs, blogs, and presidents, Journal of Happiness Studies (2009).
[8] R. Crane, F. Schweitzer, D. Sornette, Power law signature of media exposure in human response waiting time distributions, Physical Review E 81 (5) (2010) 056101.
[9] M. Mitrović, B. Tadić, Bloggers behavior and emergent communities in blog space, European Physical Journal B 73 (2) (2010) 293–301.
[10] M. Szell, R. Lambiotte, S. Thurner, Multirelational organization of large-scale social networks, Proceedings of the National Academy of Sciences USA 107 (31) (2010) 13636–13641.
[11] S. Gonzalez-Bailon, R.E. Banchs, A. Kaltenbrunner, Emotional reactions and the pulse of public opinion: measuring the impact of political events on the sentiment of online discussions, 2010, e-print arXiv:1009.4019.
[12] M. Mitrović, G. Paltoglou, B. Tadić, Networks and emotion-driven user communities at popular blogs, European Physical Journal B 77 (2010) 597–609.
[13] M. Mitrović, G. Paltoglou, B. Tadić, Quantitative analysis of bloggers' collective behavior powered by emotions, Journal of Statistical Mechanics: Theory and Experiment 2011 (02) (2011) P02005.
[14] P.S. Dodds, K.D. Harris, I.M. Koloumann, C.A. Bliss, C.M. Danforth, Temporal patterns of happiness and information in a global social network: hedonometric and twitter, 2011 arXiv:1101.5120v3.
[15] G. Paltoglou, S. Gobron, M. Skowron, M. Thelwall, D. Thalmann, Sentiment analysis of informal textual communication in cyberspace, in: Proceedings of the ENGAGE 2010 (Springer LNCS State-of-the-Art Survey), Springer, Heidelberg, 2010, pp. 13–23.
[16] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, Journal of the American Society for Information Science and Technology 61 (2010) 2544–2558.
[17] K. Scherer, What are emotions? and how can they be measured? Social Science Information 44 (4) (2005) 695–729.
[18] J.A. Russell, A circumplex model of affect, Journal of Personality and Social Psychology 39 (1980) 1161–1178.
[19] The Handbook of Emotion Elicitation and Assessment, Oxford University Press Series in Affective Science, 2007.
[20] R.A. Calvo, S. D'Mello, Affect detection: an interdisciplinary review of models, methods, and their applications, IEEE Transactions on Affective Computing 1 (1) (2010) 18–37.
[21] T. Yarkoni, Personality in 100.000 words: a large-scale analysis of personality and word use among bloggers, Journal of Research in Personality 44 (2010) 363–373.
[22] M.A. Nicolaou, H. Gunes, M. Pantic, Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space, IEEE Transactions on Affective Computing 2 (2011) 92–105.
[23] G.N. Gilbert, Agent-Based Models, Sage Publications, Los Angeles, 2008.
[24] F. Schweitzer, D. Garcia, An agent-based model of collective emotions in online communities, European Physical Journal B 77 (4) (2010) 533–545.
[25] L. Feldman Barrett, J.A. Russell, Independence and bipolarity in the structure of current affect, Journal of Personality and Social Psychology 74 (4) (1998) 967–984.
[26] M. Mitrović, B. Tadić, Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities, Physical Review E 80 (2) (2009) 026123.
[27] A. Vázquez, J.G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, A.-L. Barabási, Modeling bursts and heavy tails in human dynamics, Physical Review E 73 (3) (2006) 036127.
[28] M. Šuvakov, M. Mitrović, V. Gligorijević, B. Tadić, How online social networks are used: dialogs-based structure of MySpace, 2012, Preprint (submitted for publication).
[29] M. Mitrović, B. Tadić, Emergence and structure of cybercommunities, in: M.T. Thai, P.M. Pardalos (Eds.), Handbook of Optimization in Complex Networks, in: Springer Optimization and Its Applications, vol. 57, Springer, Berlin, 2012, pp. 209–227. http://dx.doi.org/10.1007/978-1-4614-0754-6-8. Part 2.
[30] M. Mitrović, B. Tadić, Patterns of emotional blogging and emergence of communities, Agent-Based Model on Bipartite Networks (2011) e-print arXiv:1110.5057.