PageRank model of opinion formation on social networks

Vivek Kandiah, Dima L. Shepelyansky *

Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, F-31062 Toulouse, France

ARTICLE INFO

ABSTRACT

We propose the PageRank model of opinion formation and investigate its rich properties on real directed networks of the Universities of Cambridge and Oxford, LiveJournal, and Twitter. In this model, the opinion formation of linked electors is weighted with their PageRank probability. Such a probability is used by the Google search engine for ranking of web pages. We find that the society elite, corresponding to the top PageRank nodes, can impose its opinion on a significant fraction of the society. However, for a homogeneous distribution of two opinions, there exists a bistability range of opinions which depends on a conformist parameter characterizing the opinion formation. We find that the LiveJournal and Twitter networks have a stronger tendency to a totalitarian opinion formation than the university networks. We also analyze the Sznajd model generalized for scale-free networks with the weighted PageRank vote of electors.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

To understand the nature and origins of mass opinion formation is an outstanding challenge of democratic societies [1]. In the last few years the enormous development of such social networks as LiveJournal [2], Facebook [3], Twitter [4], and VKONTAKTE [5], with up to hundreds of millions of users, has demonstrated the growing influence of these networks on social and political life. The small-world scale-free structure of the social networks (see, e.g., Refs. [6,7]), combined with their rapid communication facilities, leads to a very fast information propagation over networks of electors, consumers, and citizens, making them very active on instantaneous social events. This invokes the need for new theoretical models which would allow one to understand the opinion formation process in modern society in the 21st century.

The important steps in the analysis of opinion formation have been done with the development of various voter models, described in great detail in Refs. [8–15]. This research field became known as sociophysics [8,10,12]. In this work, we introduce several new aspects which take into account the generic features of social networks. First, we analyze the opinion formation on real directed networks taken from the Academic Web Link Database Project of British university networks [16], the LiveJournal database [17], and the Twitter dataset [18]. This allows us to incorporate the correct scale-free network structure instead of unrealistic regular lattice networks, often considered in voter models [13,14]. Second, we assume that the opinion at a given node is formed by the opinions of its linked neighbors weighted with the PageRank probability of these network nodes. We argue that the introduction of such a weight represents the reality of social networks: all the network nodes are characterized by the PageRank vector which gives the probability of finding a random surfer on a given node, as described in Refs. [19,20]. This vector gives a steady-state probability distribution on the network which provides a natural ranking of node importance, or elector or society member importance. The PageRank vector is the right eigenvector with unit eigenvalue of the Google matrix constructed from the adjacency matrix of a given directed network. A detailed

---

* Corresponding author. Tel.: +33 561556068; fax: +33 561556065.
E-mail address: dima@irsamc.ups-tlse.fr (D.L. Shepelyansky).
URL: http://www.quantware.ups-tlse.fr/dima (D.L. Shepelyansky).

description of this vector and of Google matrix construction is given in Ref. [20]. The PageRank vector is used by the Google search engine for an efficient ranking of web pages [19,20].

In a certain sense, the top nodes of PageRank correspond to a political elite of the social network whose opinion influences the opinions of other members of the society [1]. Thus the proposed PageRank opinion formation (PROF) model takes into account the situation in which an opinion of an influential friend from high ranks of the society counts more than an opinion of a friend from a lower society level. We argue that the PageRank probability is the most natural form of ranking of society members. Indeed, the efficiency of PageRank rating is demonstrated for various types of scale-free network, including the World Wide Web (WWW) [19,20], *Physical Review* citation network [21,22], scientific journal rating [23], ranking of tennis players [24], Wikipedia articles [25], the world trade network [26], and others. Due to the above argument, we consider that the PROF model captures the reality of social networks, and below we present an analysis of its interesting properties.

We note that social networks have typical features which also appear in various sciences, including the economy [27,28], trader markets [29], world trade [26], and epidemic propagation [30,31], and hence we hope that the results presented in this work will find a broad field of applications there.

The paper is composed as follows. The PROF model is described in Section 2, and the numerical results on its properties are presented in Section 3 for British university networks. In Section 4, we combine the PROF model with the Sznajd model [13,32] and study the properties of the PROF–Sznajd model. In Section 5, we analyze the models on an example of a large social network, namely LiveJournal [17]. The results for the Twitter dataset [18] are presented in Section 6. A discussion of the results is presented in Section 7.

2. PageRank opinion formation (PROF) model description

The PROF model is defined in the following way. In agreement with the standard PageRank algorithm [20], we determine the PageRank probability P_i for each node i and arrange all N nodes in monotonic decreasing order of the probability. In this way each node i has a probability $P(K_i)$, and the PageRank index K_i with the maximal probability is at $K_i = 1$ ($\sum_{i=1}^{N} P(K_i) = 1$). We use the usual damping factor value $\alpha = 0.85$ to compute the PageRank vector of the Google matrix of the network (see, e.g., Refs. [19,20,33,34]). In addition, a network node i is characterized by an Ising spin variable σ_i which can take values $+1$ or -1, coded also by red or blue color, respectively. The sign of a node i is determined by its direct neighbors j, which have PageRank probabilities P_j. For that we compute the sum Σ_i over all directly linked neighbors j of node i:

$$\Sigma_i = a \sum_j P_{j,in}^+ + b \sum_j P_{j,out}^+ - a \sum_j P_{j,in}^- - b \sum_j P_{j,out}^-, \quad a + b = 1, \tag{1}$$

where $P_{j,in}$ and $P_{j,out}$ denote the PageRank probability P_j of a node j pointing to node i (incoming link) and a node j to which node i points to (outgoing link), respectively. Here, the two parameters a and b are used to tune the importance of incoming and outgoing links with the imposed relation $a + b = 1$ ($0 \leq a, b \leq 1$). The values P^+ and P^- correspond to red and blue nodes, respectively. The spin σ_i takes the value 1 or -1, respectively, for $\Sigma_i > 0$ or $\Sigma_i < 0$. In a certain sense we can say that a large value of parameter b corresponds to a conformist society in which an elector i takes an opinion of other electors to which he/she points (nodes with many incoming links are on average at the top positions of PageRank). In contrast, a large value of a corresponds to a tenacious society in which an elector i takes mainly the opinion of those electors who point to him/her.

The condition (1) on spin inversion can be written via the effective Ising Hamiltonian H of the whole system of interacting spins:

$$H = -\sum_{i,j} J_{ij}\sigma_i\sigma_j = -\sum_i B_i\sigma_i = \sum_i \epsilon_i, \tag{2}$$

where the spin–spin interaction J_{ij} determines the local magnetic field B_i on a given node i:

$$B_i = \sum_j (aP_{j,in} + bP_{j,out})\sigma_j, \tag{3}$$

which gives the local spin energy $\epsilon_i = -B_i\sigma_i$. According to (2) and (3), the interaction between a selected spin i and its neighbors j is given by the PageRank probability: $J_{ij} = aP_{j,in} + bP_{j,out}$. Thus from a physical viewpoint the whole system can be viewed as a disordered ferromagnet [12,14]. In this way, condition (1) corresponds to a local energy ϵ_i minimization done at zero temperature. We note that such an analogy with spin systems is well known for opinion formation models on regular lattices [12–14]. However, it should be noted that generally we have asymmetric couplings $J_{ij} \neq J_{ji}$, which is unusual for physical problems (see the discussion in Ref. [35]). In view of this analogy, it is possible to introduce a finite temperature T and then to make a probabilistic Metropolis-type condition [36] for the spin i inversion determined by a thermal probability $\rho_i = \exp(-\Delta\epsilon_i/T)$, where $\Delta\epsilon_i$ is the energy difference between on-site energies ϵ_i with spin up and down. During the relaxation process, each spin is tested on an inversion condition that requires N steps and then we do t iterations of N such steps. We discuss the results of the relaxation process at both zero temperature and at finite temperature T in the next section. We use a standard random number generator to create an initial random distribution of spins σ_i up and down on
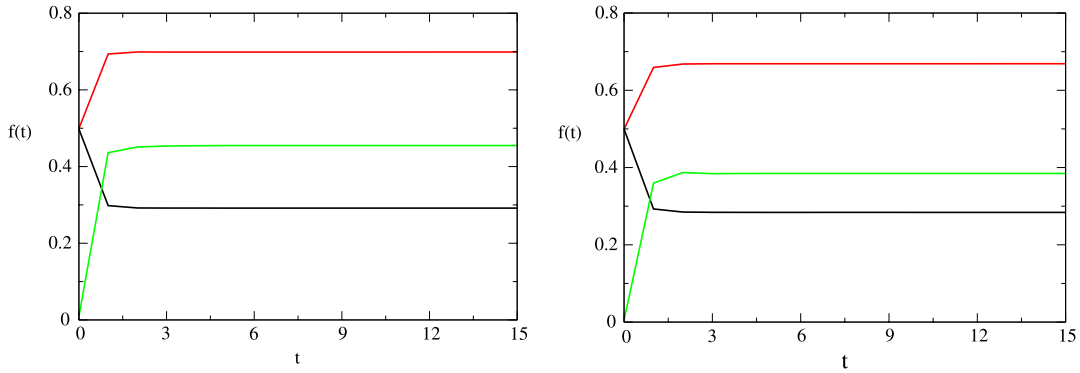
Fig. 1. (Color online) Time evolution of opinion given by a fraction of red nodes $f(t)$ as a function of number of iterations t. The red and black curves (top and bottom curves at $t = 15$, respectively) show evolution for two different realizations of a random distribution of color with the same initial fraction $f_i = 0.5$ at $t = 0$. The green curve (middle curve at $t = 15$) shows the dependence $f(t)$ for the initial state with N_{top} all red nodes with top PageRank K indexes (highest $P(K_i)$ values, $1 \le K \le N_{top}$). The evolution is done at $a = b = 0.5$ and temperature $T = 0$. *Left panel:* Cambridge network with $N_{top} = 2000$. *Right panel:* Oxford network with $N_{top} = 1000$.

nodes of a given network. We do averaging over $N_r \le 10^4$ such random generations to obtain statistically stable results for the final opinion distributions. The Metropolis Monte Carlo simulations follow the standard procedure described in Ref. [36].

3. Numerical results for the PROF model on university networks

Here we present results for the PROF model considered on the networks of the Universities of Cambridge and Oxford in 2006, taken from Ref. [16]. The properties of PageRank distribution $P(K)$ for these networks have been analyzed in Refs. [33,34]. The total numbers of nodes N and links N_ℓ are $N = 212710$, $N_\ell = 2015265$ (Cambridge); and $N = 200823$, $N_\ell = 1831542$ (Oxford) [34]. Both networks are characterized by an algebraic decay of PageRank probability $P(K) \propto 1/K^\beta$ and approximately usual exponent value $\beta \approx 0.9$; additional results on the scale-free properties of these networks are given in Refs. [33,34]. We usually discuss the fraction of red nodes, since by definition all other nodes are blue.

Typical examples of time evolution of the fraction of red nodes $f(t)$ with the number of time iterations t are shown in Fig. 1. We see the presence of bistability in the opinion formation: two random states with the same initial fraction of red nodes $f_i = f(t = 0)$ evolve to two different final fractions of red nodes f_f. The process gives an impression of convergence to a fixed state after $t_c \approx 10$ iterations. A special check shows that all node colors become fixed after this time (t_c). The convergence time to a fixed state is similar to those found for opinion formation on regular lattices, where $t_c = O(1)$ [13,14,37]. The corresponding time evolution of colors is shown in Fig. 2 for the first 10% of nodes ordered by the PageRank index K.

The results of Fig. 1 show that for a random initial distribution of colors we may have different final states with ± 0.2 variation compared to the initial $f_i = 0.5$. However, if we consider that N_{top} nodes with the top K index values (from 1 to N_{top}) have the same opinion (e.g. red nodes), then we find that even a small fraction of the total number of nodes N (e.g. N_{top} of about 0.5% or 1% of N) can impose its opinion on a significant fraction of nodes of about $f_f \approx 0.4$. This shows that in the frame of PROF model the society elite, corresponding to the top K nodes, can significantly influence the opinion of the whole society under the condition that the elite members have a fixed opinion between themselves.

We also considered the case when the red nodes are placed on $N_{top} = 2000$ top nodes of the CheiRank index K^*. This ranking is characterized by the CheiRank probability $P^*(K^*)$ for a random surfer moving in the inverted direction of links, as described in Refs. [25,34]. On average $P^*(K^*)$ is proportional to the number of outgoing links. However, in this case, the top nodes with small f_i values are not able to impose their opinion, and the final fraction becomes blue. We attribute this to the fact that the opinion condition (1) is determined by the PageRank probability $P(K)$ and that the correlations between CheiRank and PageRank are not very strong (see the discussion in Refs. [25,34]).

To analyze how the final fraction of red nodes f_f depends on its initial fraction f_i, we study the time evolution $f(t)$ for a large number N_r of initial random realizations of colors following it up to the convergence time for each realization. We find that the final red nodes are homogeneously distributed in K. Thus there is no specific preference for top society levels for an initial random distribution. The probability distribution W_f of final fractions f_f is shown in Fig. 3 as a function of initial fraction f_i at three values of parameter a. These results show two main features of the model: a small fraction of red opinion is completely suppressed if $f_i < f_c$ and its larger fraction dominates completely for $f_i > 1 - f_c$; there is a bistability phase for the initial opinion range $f_b \le f_i \le 1 - f_b$. Of course, there is a symmetry in respect to exchange of red and blue colors. For the small value $a = 0.1$ we have $f_b \approx f_c$ with $f_c \approx 0.25$, while for the large value $a = 0.9$ we have $f_c \approx 35$, $f_b \approx 0.45$.

Our interpretation of these results is the following. For small values of $a \to 0$ the opinion of a given society member is determined mainly by the PageRank of neighbors *to whom he/she points* (outgoing links). The PageRank probability P of nodes to which many nodes point is usually high, since P is proportional to the number of ingoing links [20]. Thus, at
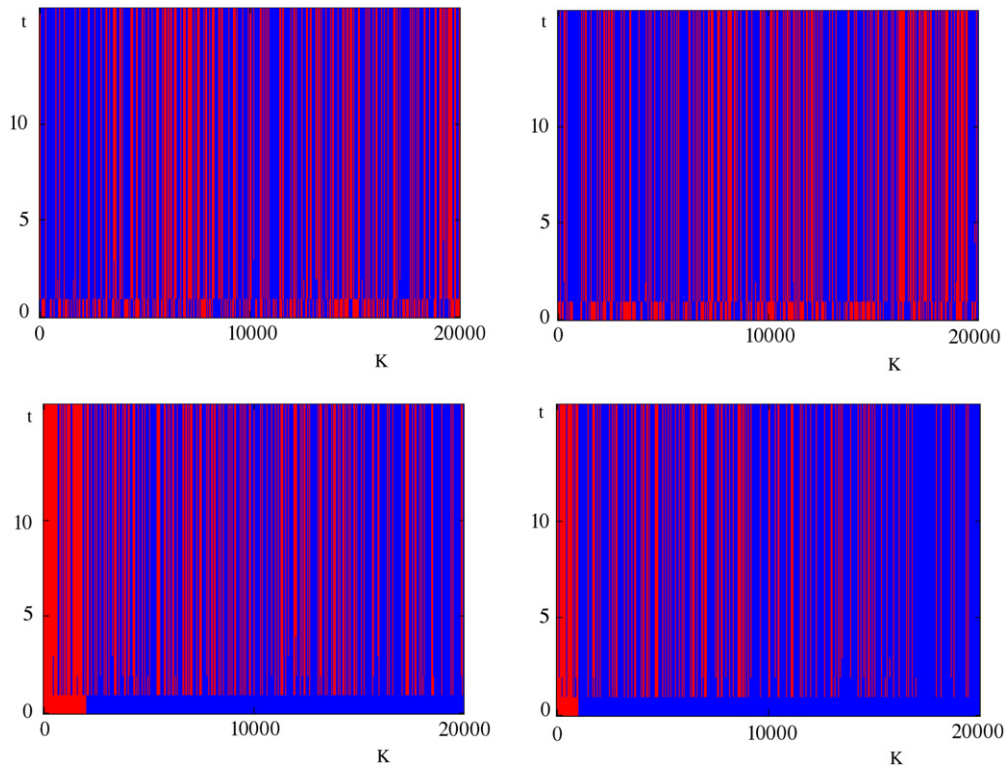
Fig. 2. (Color online) Time evolution of opinion colors (red/gray and blue/black) for the parameters of Fig. 1: the left/right column is for the Cambridge/Oxford network. The initial fraction of red colors is $f_i = 0.5$ (top panel), and N_{top} nodes have red color for the bottom panels, with $N_{top} = 2000$ and 1000 for the Cambridge network and the Oxford network, respectively. Nodes are ordered by the PageRank index K, and the color plot shows only $K \leq 20000$.

$a \rightarrow 0$, the society is composed of members who form their opinion by listening to an elite opinion. In such a society its elite with one color opinion can impose this opinion on a large fraction of the society. This is illustrated in Fig. 4, which shows a dependence of the final fraction f_f of red nodes on parameter a for a small initial fraction of red nodes in the top values of the PageRank index ($N_{top} = 2000$). We see that $a = 0$ corresponds to a conformist society which follows in its great majority the opinion of its elite. For $a = 1$, this fraction f_f drops significantly, showing that this corresponds to a regime of a tenacious society. It is somewhat surprising that the tenacious society ($a \rightarrow 1$) has a well-defined and relatively large fixed opinion phase with a relatively small region of bistability phase. This is in contrast to the conformist society at $a \rightarrow 0$, where the opinion is strongly influenced by the society elite. We attribute this to the fact that in Fig. 3 we start with a randomly distributed opinion, because the opinion of the elite has two fractions of two colors that create a bistable situation, since the two fractions of society follow the opinions of this divided elite, which makes the situation bistable on a larger interval of f_i compared to the case of a tenacious society at $a \rightarrow 1$.

To stress the important role of PageRank in the dependence of f_f on f_i presented in Fig. 3, we show in Fig. 5 the same analysis at $a = 0.5$, but for the case when in Eq. (1) for the spin flip we take all $P = 1$ (equal weight for all nodes). The data of Fig. 5 clearly demonstrate that in this case the bistability of opinion disappears. Thus the PROF model is qualitatively different from the case when only the links without their PageRank weight are counted for the spin flip condition. We also test the sensitivity in respect to PageRank probability by replacing P by $\sqrt{P}$ in Eq. (1), as is shown in Fig. 5 (bottom panels). We see that compared to the case $P = 1$ we start to have some signs of bistability, but still they remain rather weak compared to the case of Fig. 3.

In fact the spin flip condition (1) can be viewed as a relaxation process in a disordered ferromagnet (since all $J_{ij} \geq 0$ in (2) and (3)) at zero temperature. Such a type of analysis of voter model relaxation processes on regular lattices is analyzed in Refs. [13,14]. From this viewpoint it is natural to consider the effect of finite temperature T on this relaxation. At finite T, the flip condition is determined by the thermal Metropolis probability $\exp(-\Delta\epsilon_i/T)$, as described above. We follow this thermodynamic relaxation process at finite temperature up to $t = 200$ iterations, and in this way obtain the probability distribution of the final fraction f_f of red nodes obtained from the initial fraction f_i of red nodes randomly distributed over the network at $t = 0$. The results obtained at finite temperatures are shown in Fig. 6. They show that a finite temperature T allows a finite fraction f_f of red nodes when for their small initial fraction f_i all final f_f were equal to zero. Also, the bistability splitting is reduced and it disappears at larger values of T. Thus finite T introduces a certain smoothing in the W_f distribution.
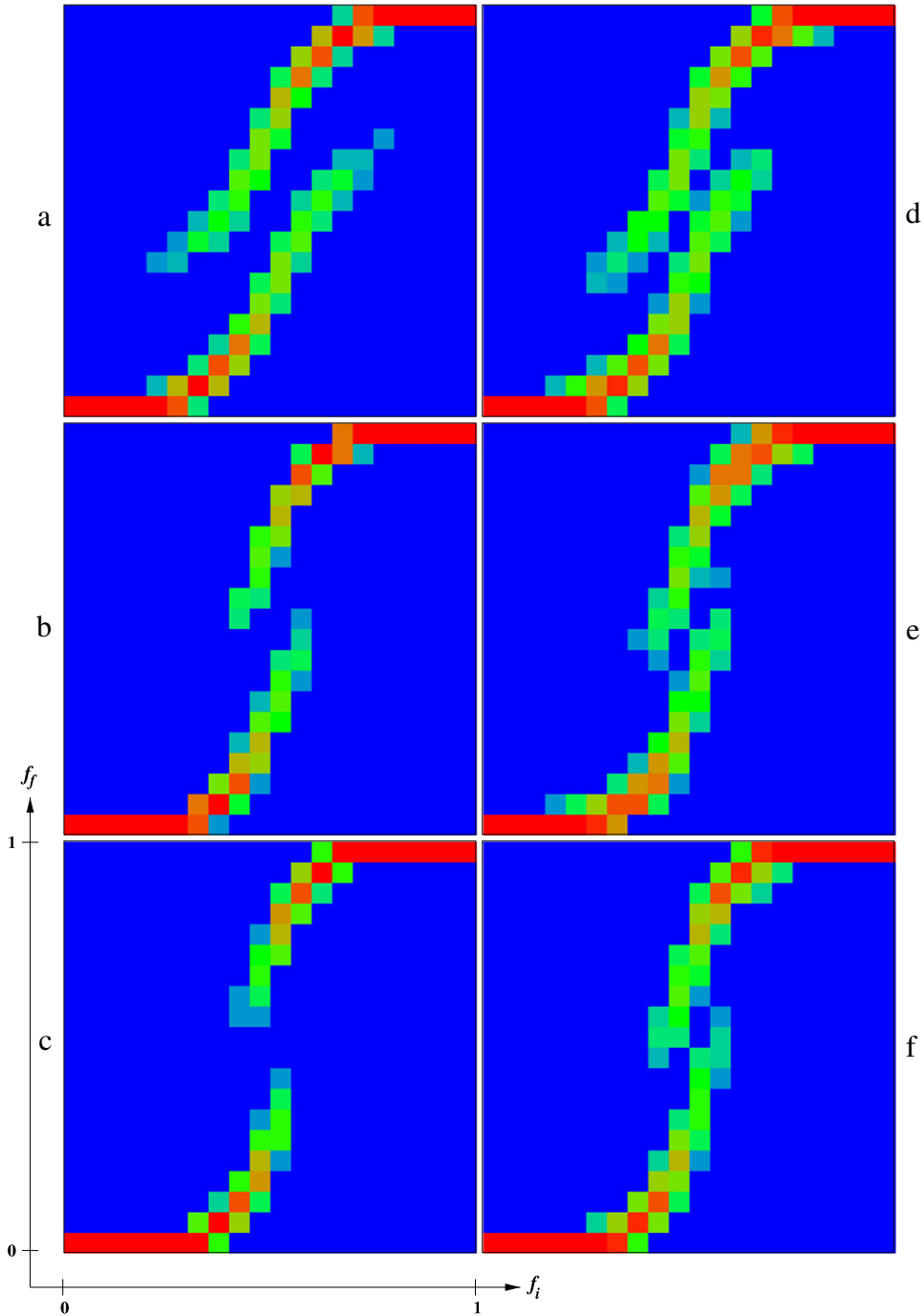
Fig. 3. (Color online) Density plot of probability W_f to find the dependence of the final red fraction f_f, shown on the y-axis, on the initial red fraction f_i, shown on the x-axis; data are shown inside the unit square $0 \le f_i, f_f \le 1$. The values of W_f are defined as the relative number of realizations found inside each of 20×20 cells which cover the whole unit square. Here, $N_r = 10^4$ realizations of randomly distributed colors are used to obtain the W_f values; for each realization, the time evolution is followed up to the convergence time with up to $t = 20$ iterations; here $T = 0$. *Left column:* Cambridge network (a, b, c); *right column:* Oxford network (d, e, f); here, $a = 0.1$ (a, d), 0.5 (b, e), 0.9 (c, f) from top to bottom. The probability W_f is proportional to color changing from zero (blue/black) to unity (red/gray).

However, the relaxation process at finite temperatures does not lead to the thermal Boltzmann distribution. Indeed, in Fig. 7, we show the probability distribution $w_i(\epsilon_i)$ as a function of local energies ϵ_i defined in (2) and (3). The distribution $w_i(\epsilon_i)$ is obtained from the relaxation process with many initial random spin realizations N_r. Even if the temperature T is
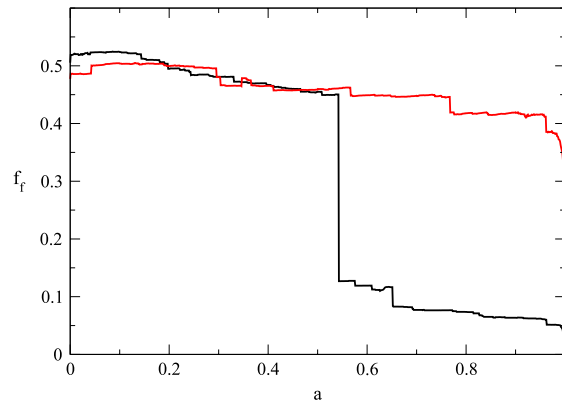
Fig. 4. (Color online) Dependence of the final fraction of red nodes f_f on the tenacious parameter a (or conformist parameter $b = 1 - a$) for initial red nodes in $N_{top} = 2000$ values of the PageRank index ($1 \leq K \leq N_{top}$); black and red/gray curves show data for Cambridge and Oxford networks; here, $T = 0$.
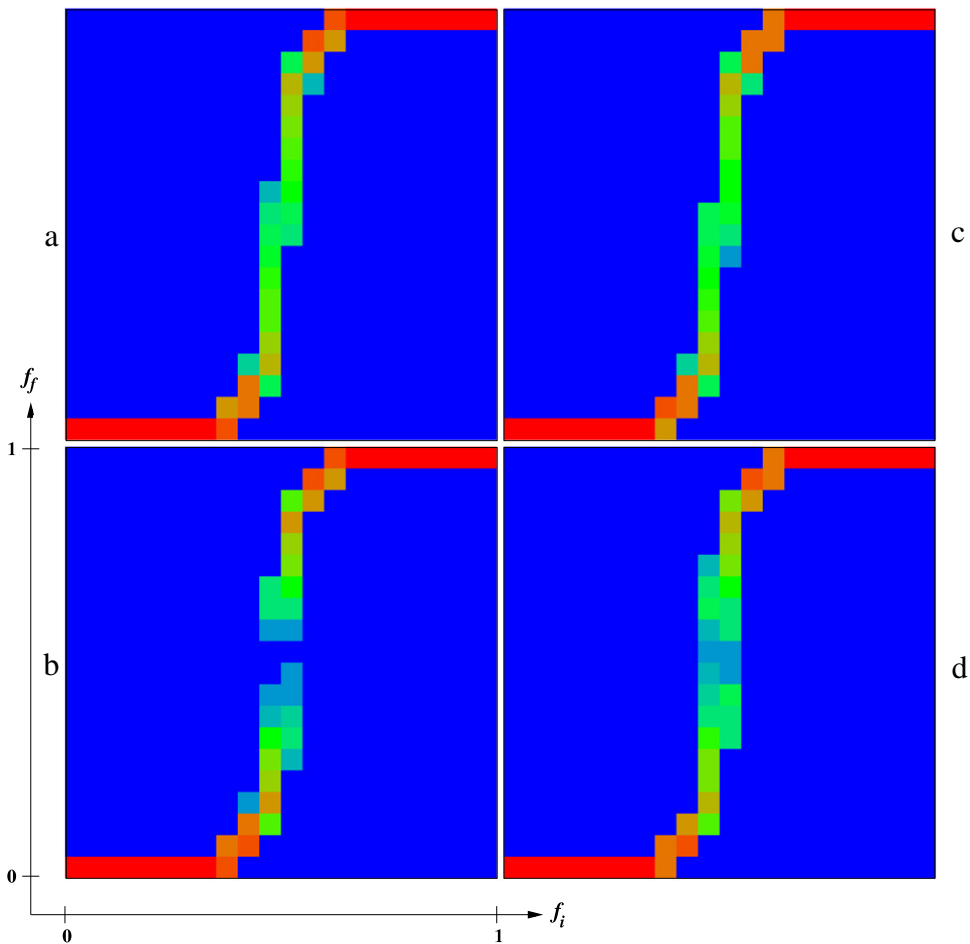

Fig. 5. (Color online) The same as in Fig. 3 (middle panels) at $a = 0.5$ but with uniform condition for spin flip being independent of PageRank probability (top panels (a, c): $P = 1$ in Eq. (1)) and PageRank probability P replaced by $\sqrt{P}$ in Eq. (1) (bottom panels (b, d)); the left and right panels correspond to Cambridge (a, b) and Oxford (c, d) networks; here, $T = 0$, and $N_r = 10^4$ realizations are used.

comparable with typical values of local energies ϵ_i, we still obtain a rather peaked distribution at $\epsilon_i \approx 0$ being very different from the Boltzmann distribution.

We argue that a physical reason of significantly non-Boltzmann distribution is related to the local nature of the spin flip condition which does not allow the production of a good thermalization on the scale of the whole system. Indeed, there are various energetic branches, and probably nonlocal thermalization flips of group of spins are required for a better
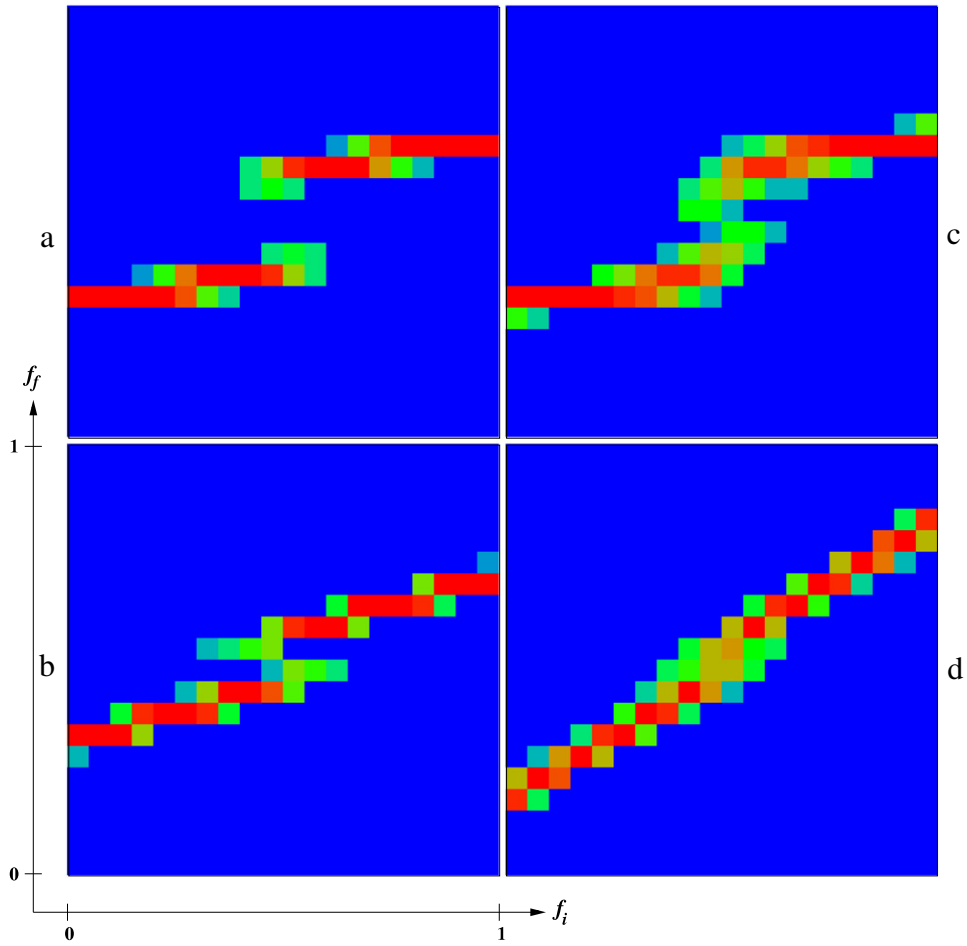
Fig. 6. (Color online) The same as in Fig. 3 (middle panel) at $a = 0.5$, but at finite temperature T during the relaxation process with $T = 0.001$ (top panels (a, c)) and $T = 0.01$ (bottom panels (b, d)); the number of random initial realizations is $N_r = 6000$, and the relaxation is done during $t = 200$ iterations. Left and right columns correspond to Cambridge (a, b) and Oxford (c, d) networks.
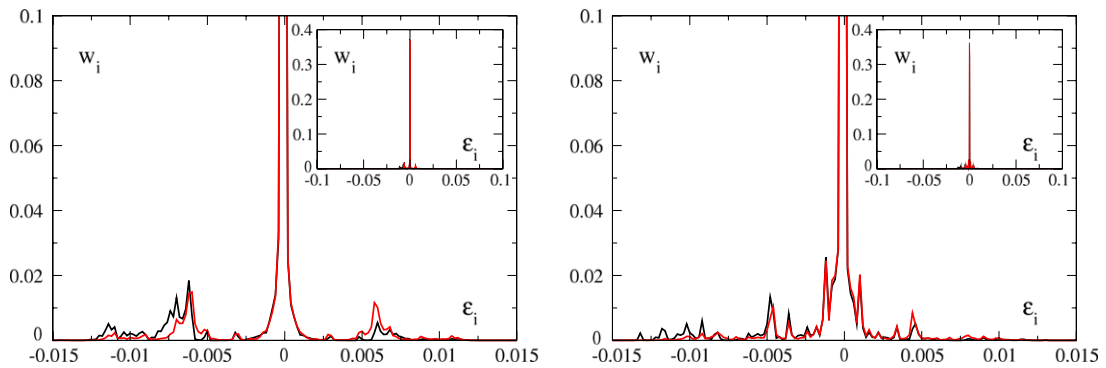
Fig. 7. (Color online) Normalized histograms of probability distribution w_i over local energies ϵ_i obtained from the relaxation process during $t = 10^3$ time iterations at temperatures $T = 0.01$ (black curve) and $T = 0.05$ (red/gray curve); the average is taken over $N_r = 200$ random initial realizations. The insets show the distributions on a large scale including all local energies ϵ_i. The left and right panels show Cambridge and Oxford networks.

thermalization. However, voting is a local process that involves only direct neighbors, which seems to be not sufficient for the emergence of a global thermal distribution. The presence of a few energy branches is well visible from the data of Fig. 8 obtained at $T = 0$. This figure shows the dependence of the final fraction f_f of red nodes on their initial fraction f_i and the total initial energy $E_i = \sum_{m=1}^{N} \epsilon_m$ of the whole system corresponding to a chosen initial random configuration of spins. Most probably, these different branches prevent efficient thermalization of the system with only local spin flip procedure.
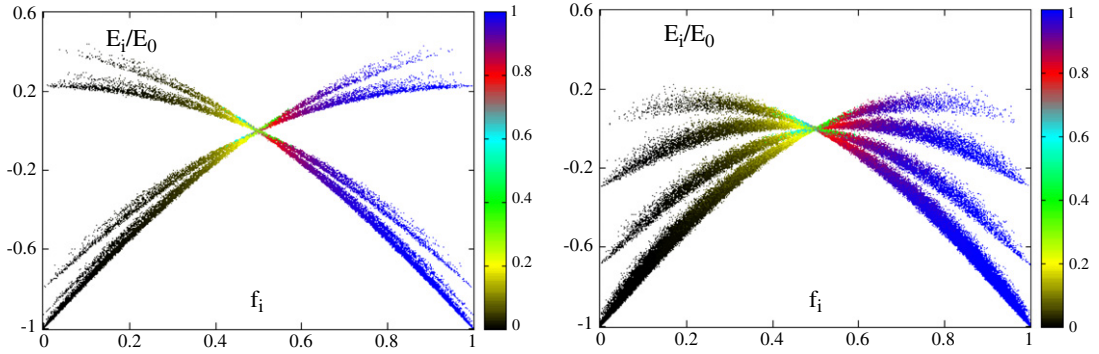
Fig. 8. (Color online) This diagram shows the final fraction of red nodes f_f, coded by color from $f_f = 0$ (black) to $f_f = 1$ (blue/dark gray), as a function of initial fraction of red nodes f_i and the total initial energy E_i; each of $N_r > 3.5 \times 10^4$ random realizations is shown by color point; data are shown after $t = 20$ time iterations at $T = 0$. The energy E_0 is the modulus of total energy with all spin up; here, $a = 0.5$. Left and right panels show data for Cambridge ($E_0 = 341.20$) and Oxford ($E_0 = 254.28$) networks; bars show color attribution to final probability f_f.

In addition to the above points, the asymmetric form of J_{ij} couplings plays an important role, generating a more complicated picture compared to the usual image of thermal relaxation (see, e.g., Ref. [35]). We also note that thermalization is absent in voter models on regular lattices [13].

4. PROF–Sznajd model

The Sznajd model [32] nicely incorporates the well-known trade union principle "United we stand, divided we fall" into the field of voter modeling and opinion formation on regular networks. A review of various aspects of this model is given in Ref. [13]. Here, we generalize the Sznajd model to include in it the features of the PROF model, and consider it on social networks with their scale-free structure. This gives us the PROF–Sznajd model, which is constructed in the following way. For a given network, we determine the PageRank probability $P(K_i)$ and the PageRank index K_i for all i nodes. We introduce the definition of a *group* of nodes. A group of nodes is defined by the following rule applied at each time step τ.

- (i) Pick by random a node i in the network and consider the polarization of the $N_g - 1$ highest PageRank nodes pointing to it.
- (ii) If node i and all other $N_g - 1$ nodes have the same color (same spin polarization), then these N_g nodes form a group whose effective PageRank value is the sum of all the member values $P_g = \sum_{j=1}^{N_g} P_j$; if this is not the case, then we leave the nodes unchanged and perform the next time step.
- (iii) Consider all the nodes *pointing to any member of the group* (this corresponds to model *option* 1) or consider *all the nodes pointing to any member of the group and all the nodes pointed by any member of the group* (this corresponds to model *option* 2); then check all these nodes n directly linked to the group: if an individual node PageRank value P_n is less than P_{group} then this node joins the group by taking the same color (polarization) as the group nodes; if this is not the case, then the node is left unchanged; the PageRank values of added nodes are then added to the group PageRank P_{group} and the group size is increased.

The above time step is repeated many times during time τ, counting the number of steps, by choosing a random node i on each next step. This procedure effectively corresponds to the zero-temperature case in the PROF model.

A typical example of the time evolution of the fraction of red nodes $f(\tau)$ in the PROF–Sznajd model is shown in Fig. 9. It shows that the system converges to a steady state after a time scale $\tau_c \approx 10N$ that is comparable with the convergence times for the PROF models studied in previous sections. We see that there are still some fluctuations in the steady-state regime which are visibly smaller for the option 2 case. We attribute this to a larger number of direct links in this case. The number of group nodes N_g gives some variation of f_f, but these variations remain on a relatively small scale of a few percent. Here, we should point on the important difference between the PROF and PROF–Sznajd models: for a given initial color realization, in the first case we have convergence to a fixed state after some convergence time, while in the second case we have convergence to a steady state which continues to fluctuate in time, keeping the color distribution only on average.

The dependence of the final fraction of red nodes f_f on its initial value f_i is shown by the density plot of probability W_f in Fig. 10 (option 1 of the PROF–Sznajd model). The probability W_f is obtained from many initial random realizations in a similar way to the case of Fig. 3. We see that there is a significant difference compared to the PROF model (Fig. 3): now even at small values of f_i we find small but finite values of f_f, while in the PROF model the red color disappears at $f_i < f_c$. This feature is related to the essence of the Sznajd model: here, even small groups can resist the totalitarian opinion. Other features of Fig. 10 are similar to those found for the PROF model: we again observe bistability of opinion formation. The number of nodes N_g, which form the group, does not significantly affect the distribution W_f: we have smaller fluctuations at larger N_g values but the model already works in a stable way at $N_g = 3$. The results for option 2 of the PROF–Sznajd model are shown in Fig. 11. In this case, the opinions with a small initial fraction of red nodes f_i are suppressed in a significantly
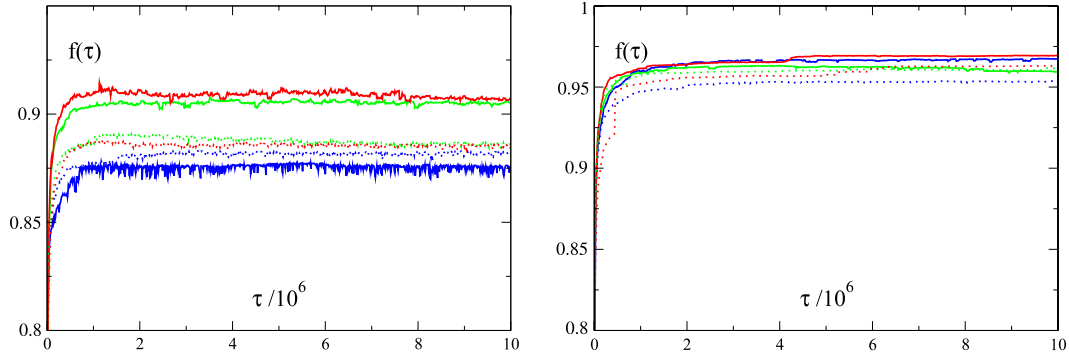
Fig. 9. (Color online) Time evolution of the fraction of red nodes $f(\tau)$ in the PROF–Sznajd model with the initial fraction of red nodes $f_i = 0.7$ at one random realization. The curves show data for three values of group size $N_g = 3$ (blue/black); 8 (green/light gray); and 13 (red/gray). Full/dashed curves are for Cambridge/Oxford networks; the left panel is for option 1; the right panel is for option 2.

stronger way compared to option 1. We attribute this to the fact that large groups can suppress small groups in a stronger way, since the outgoing direct links are taken into account in this option.

The significant difference between the two options of the PROF–Sznajd model is well seen from the data of Fig. 12. Here, all N_{top} nodes are taken in red (compare with the PROF model in Fig. 4). For option 1, the society elite succeeds in imposing its opinion on a significant fraction of nodes, which is increased by a factor 5–10. Visibly, this increase is less significant than in the PROF model. However, for option 2 of the PROF–Sznajd model there is practically no increase of the fraction of red nodes. Thus, in option 2 the society members are very independent and the influence of the elite on their opinion is very weak.

5. PROF models on the LiveJournal network

Even if one can expect that the properties of university networks are similar to those of real social networks, it is important to analyze the previous PROF models in the frame of a real social network. For that we use the LiveJournal network, collected, described, and presented in Ref. [17]. From this database we obtain a directed network with $N = 3577166$ nodes and $N_\ell = 44913072$ links, which are mainly directed (only about 30% of links are symmetric). The Google matrix of the network is constructed in the usual way [20], and its PageRank vector is determined by the iteration process at damping factor $\alpha = 0.85$. For the time evolution of fraction of red nodes f we use time iterations in t and τ defined as in previous sections.

The PageRank probability decay $P(K)$ is shown in Fig. 13. It is well described by an algebraic law $P(K) \propto 1/K^\beta$ with $\beta = 0.448 \pm 0.000046$. The convergence of a fraction of red nodes $f(t)$ takes place approximately on the same convergence time scale $t_c \sim 5 \sim O(1)$ even though the size of the network is increased almost by a factor 20.

In a way similar to the university networks we find that the homogeneous opinion of the society elite presented in a small fraction of N_{top} nodes influences a large fraction of the whole society especially when the parameter a is not very large (see Fig. 14 in comparison with Fig. 4). The influence of the elite at 1% of red nodes is larger in the case of the LiveJournal network. It is possible that this is related to a 30% larger number of links, but it is also possible that other structural network parameters also play a role here.

In spite of certain similarities with the previous data for university networks discussed before, we find that the opinion diagram for the LiveJournal network (see Fig. 14 right panel) is very different from those obtained for the university networks (see Fig. 3): the bistability has practically disappeared. We think that this difference originates from a significantly slower decay exponent for PageRank probability $P(K)$ in the case of LiveJournal. To check this assumption we compare the probability distribution W_f of final opinion f_f for an initial opinion fixed at $f_i = 0.4$ using the PROF model with the usual linear weight P in Eq. (1) and a quadratic weight proportional to P^2 (see Fig. 15). For the linear weight, we find that only very small values of $f_f \approx 0.005$ can be found for initial $f_i = 0.4$, while for the quadratic weight we obtain a rather broad distribution of f_f values in the main range $0 < f_f < 0.15$ with a few large values $f_f \approx 0.6$. Thus we see that the final opinion is rather sensitive to the weight used in Eq. (1). However, in contrast to the university networks (see Figs. 3 and 5), where we have narrow one-peak or double-peak distributions of f_f, for the LiveJournal network with quadratic weight we find a rather broad distribution of f_f. In the spirit of a renormalization map description considered in Ref. [10] (see Figs. 1, 2 there), it is possible to assume that one or two peaks corresponds to one or two fixed point attractors of the map. We make a conjecture that a broad distribution as in Fig. 15 (right panel) can correspond to a regime of a strange chaotic attractor appearing in the renormalization map dynamics. In principle, such a chaotic renormalization dynamics is known to appear in coupled spins lattices when three-spin couplings are present (see Ref. [38] and the references therein). It is possible that the presence of weight probability associated with the PageRank in a certain power may lead to chaotic dynamics which would generate a broad distribution of final opinions f_f.

We also made tests for the PROF–Sznajd model (option 1) for the LiveJournal database. However, in this case, at $f_i = 0.4$ and $a = 0.5$, we found only small f_f values (similar of those in Fig. 15, left panel) both for linear and quadratic weights in Eq. (1). It is possible that the Sznajd groups are less sensitive to the probability weight.
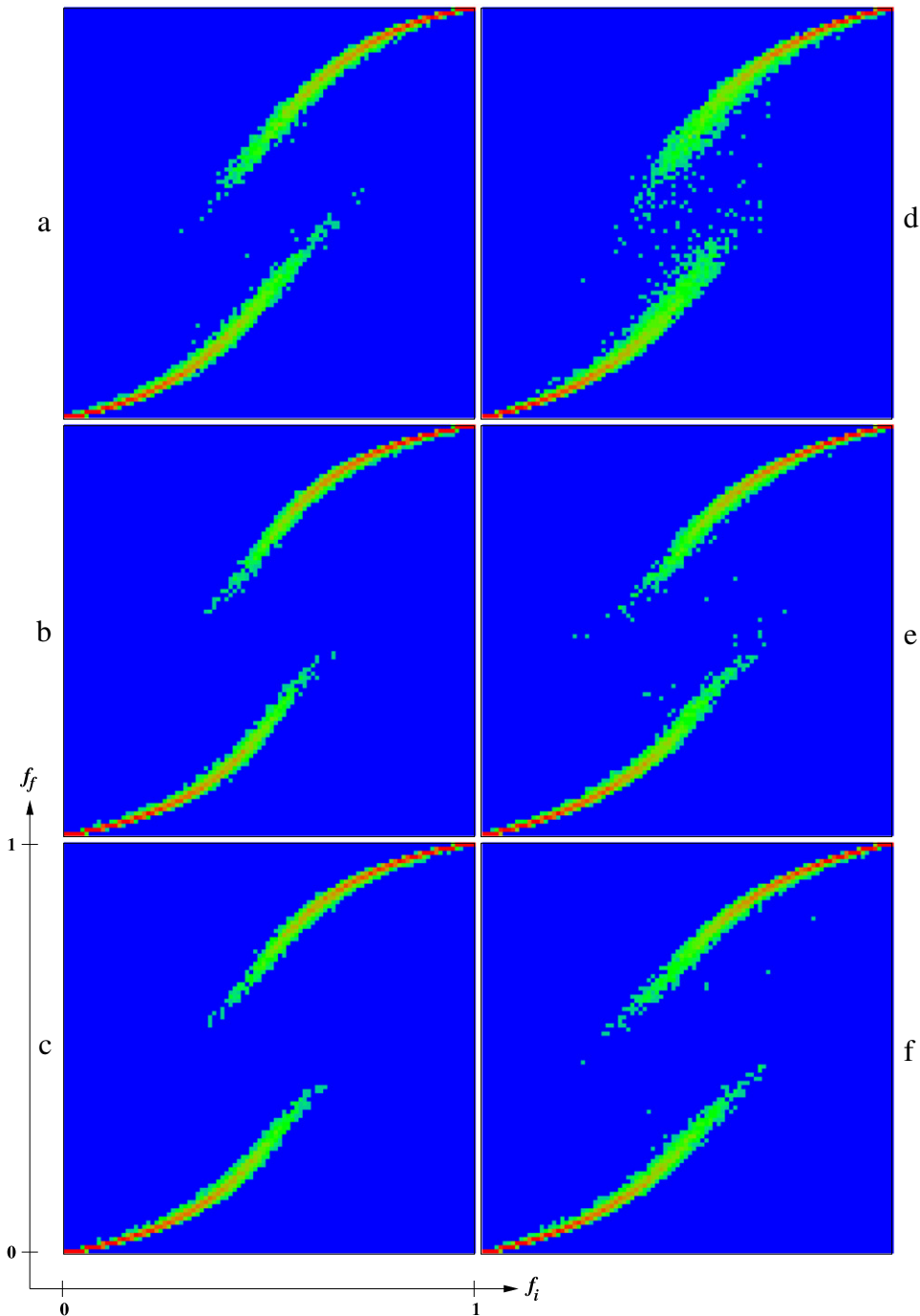
Fig. 10. (Color online) PROF–Sznajd model, option 1: density plot of probability W_f to find the dependence of the final red fraction f_f, shown on the y-axis, on the initial red fraction f_i, shown on the x-axis; data are shown inside the unit square $0 \le f_i, f_f \le 1$. The values of W_f are defined as the relative number of realizations found inside each of 100×100 cells which cover the whole unit square. Here, $N_r = 10^4$ realizations of randomly distributed colors are used to obtained W_f values; for each realization the time evolution is followed up to the convergence time with up to $\tau = 10^7$ steps. *Left column:* Cambridge network (a, b, c); *right column:* Oxford network (d, e, f); here, $N_g = 3$ (a, d), 8 (b, e), 13 (c, f) from top to bottom. The probability W_f is proportional to color changing from zero (blue/black) to unity (red/gray).

6. PROF models for the Twitter dataset

We also analyzed the opinion formation on the Twitter dataset with $N = 41\,652\,230$, $N_\ell = 1\,468\,365\,182$ taken from Ref. [18]. This is the entire size of Twitter at the corresponding moment of time [18]. The size is rather large, and due to that we present only the main features of the PROF model for this directed network.
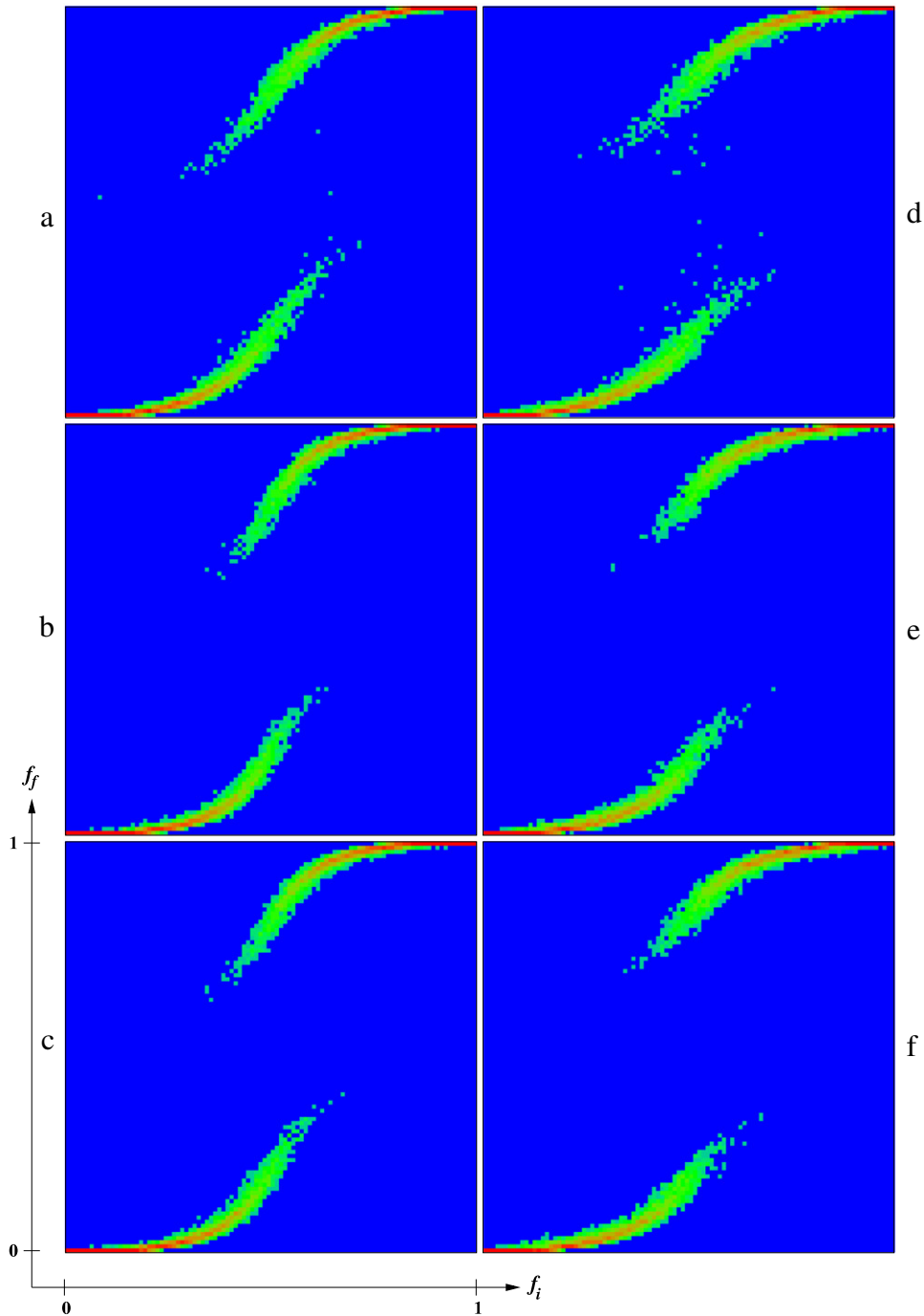
Fig. 11. (Color online) The same as in Fig. 10 but for PROF–Sznajd model, option 2.

The dependence of PageRank P on its index K is shown in Fig. 16 (left panel). For the range $1 \leq \log_{10} K \leq 5.5$, we find that the decay exponent $\beta \approx 0.51$, being similar to that of the LiveJournal network (see Fig. 13) even if there is a faster drop of P at larger K values. We note that the value $\beta \approx 0.5$ is rather different from the value usually found for the Zipf law [39] and the WWW [20], with $\beta \approx 1$. It is possible that this is related to a significantly larger average number of links per node, which is increased by a factor 3.5 for the Twitter network compared to the university networks analyzed in the previous sections.

The effect of the homogeneous elite opinion of all red N_{top} nodes is shown in Fig. 16 (right panel). We see that on the Twitter network a small fraction of elite with fixed opinion ($N_{top}/N \approx 3 \times 10^{-5}$) can impose this opinion on practically
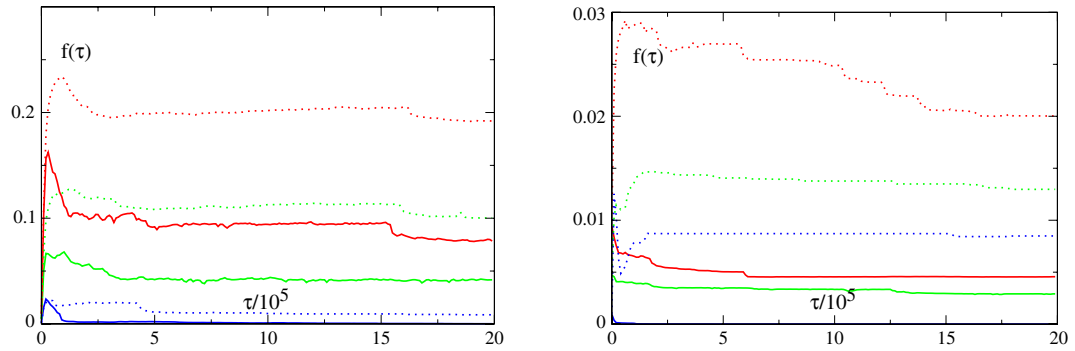
Fig. 12. (Color online) Time evolution of the fraction of red nodes $f(\tau)$ in the PROF–Sznajd model with the initial red nodes for the top PageRank nodes: $N_{top} = 200$ (blue/black); 1000 (green/light gray); 2000 (red/gray); here, $N_g = 8$. Full/dashed curves are for Cambridge/Oxford networks; the left panel is for option 1; the right panel is for option 2. The color of curves is red, green, blue, from top to bottom at maximal τ on both panels.
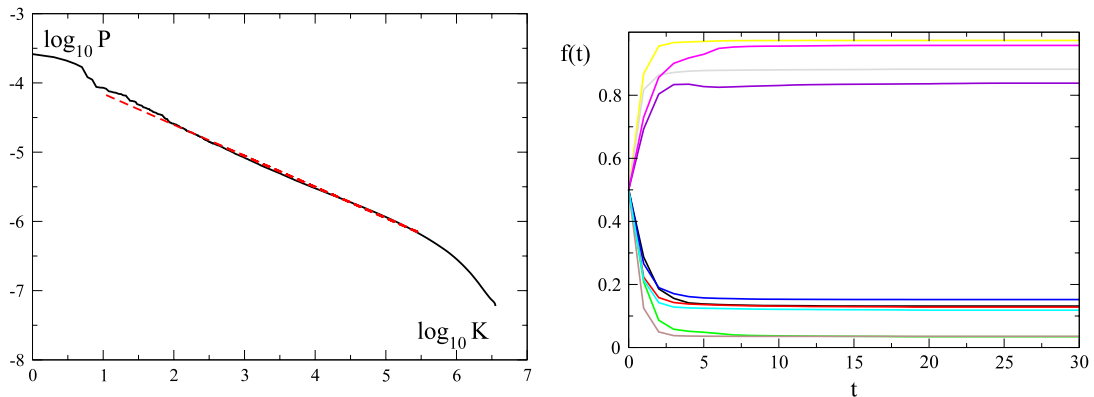
Fig. 13. (Color online) Data for the LiveJournal network. *Left panel:* PageRank probability decay with PageRank index K (full curve); the fitted algebraic dependence is shown by the dashed line $y = b - \beta x$ (for $1 \leq \log_{10} K \leq 5.5$) with the exponent $\beta = 0.448 \pm 0.000046$ and $b = -3.70 \pm 0.00023$. *Right panel:* time evolution of opinion given by a fraction of red nodes $f(t)$ as a function of number of iterations t (cf. Fig. 1) at $a = 0.5$; a few random initial realizations with $f_i = 0.5$ are shown.

the whole community for all values of the conformist parameter $1 - a$. We find that for $N_{top} > 1300$ all f_f values are very close to unity, while for $N_{top} < 1200$ we find $f_f = 0$, as is seen in Fig. 16, right panel. Thus, the transition is very sharp. We attribute such a strong influence of elite opinion to the very connected structure of Twitter network with a significantly larger average number of links per node compared to the university and LiveJournal networks.

At $a = 0.5$, for a fixed fraction of initial opinion $f_i = 0.4$, we find that the probability distribution W_f of final opinion f_f is located in the range of small values $0.0006 < W_f < 0.0007$ for both the linear P and quadratic P^2 weights used in Eq. (1) (we do not show these data). For the linear weight, the situation is rather similar to the case of LiveJournal (see Fig. 15), but for the quadratic weight we find a significant difference between the two networks (see Fig. 15). The reason for such a significant difference for the quadratic weight case requires a more detailed comparison of network properties.

The large size of the Twitter network makes numerical simulations of the PROF–Sznajd model rather difficult, and therefore we did not study this model for this network.

7. Discussion

In this work we have proposed the PageRank model of opinion formation of social networks and analyzed its properties on examples of four different networks. For two university networks we find rather similar properties of opinion formation. Opinion formation is characterized by an important feature according to which the society elite with a fixed opinion can impose its opinion on a significant fraction of the society members which is much larger than the initial elite fraction. However, when the initial opinions of society members, including the elite, are presented by two options, then we find a significant range of opinion fraction within a bistability regime. This range depends on the conformist parameter, which characterizes the local aspects of opinion formation of linked society members. The generalization of the Sznajd model for scale-free social networks gives interesting examples of opinion formation where finite small-size groups can keep their
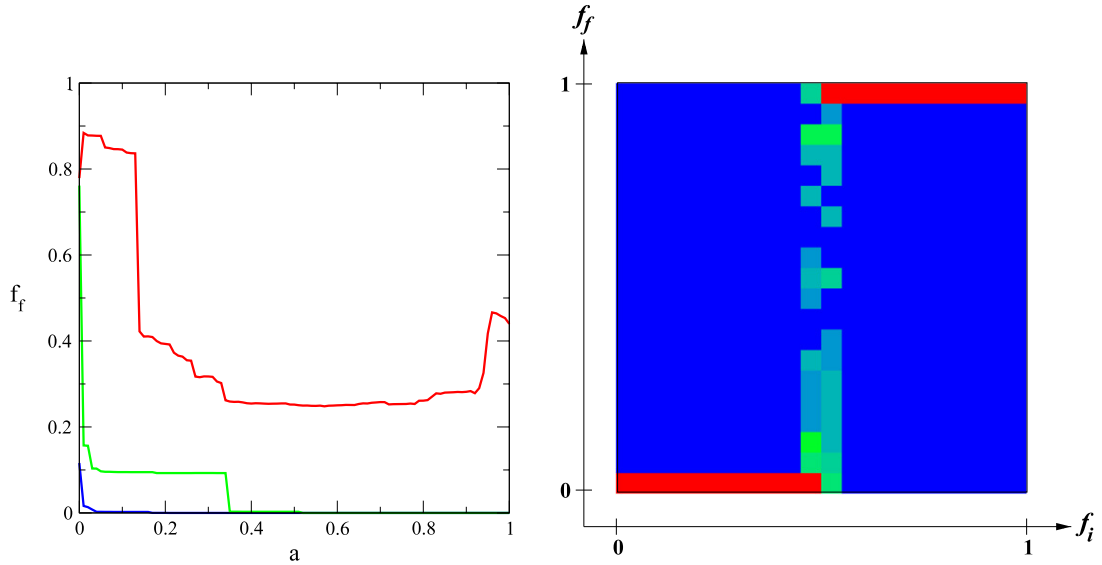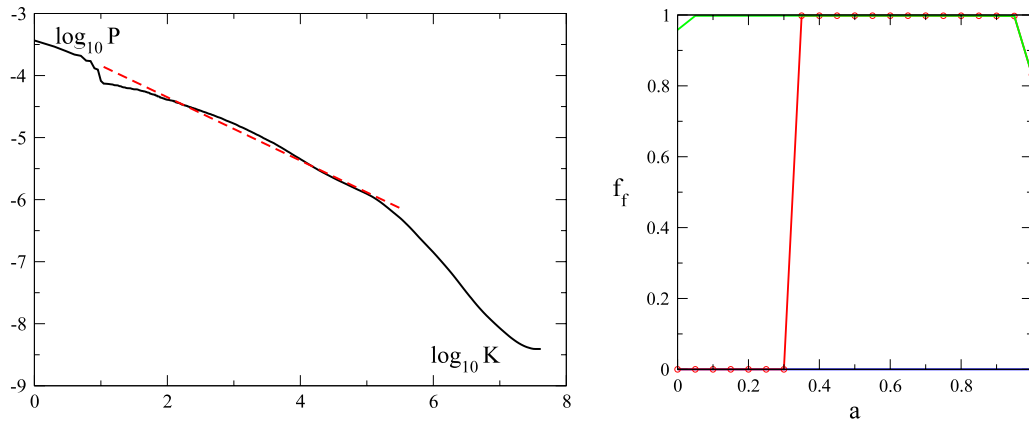
Fig. 14. (Color online) Data for the LiveJournal network. *Left panel:* dependence of the final fraction of red nodes f_f on the tenacious parameter a (or conformist parameter $b = 1 - a$) in the PROF model for initial red nodes in N_{top} values of the PageRank index ($1 \leq K \leq N_{top}$; cf. Fig. 4). Here, $N_{top} = 2000$ blue, 10,000 green, and 35,000 red curves (from bottom to top at $a = 0.5$); $T = 0$. *Right panel:* the same data as in Fig. 3 at $a = 0.5$ with the same parameters but for the LiveJournal network.
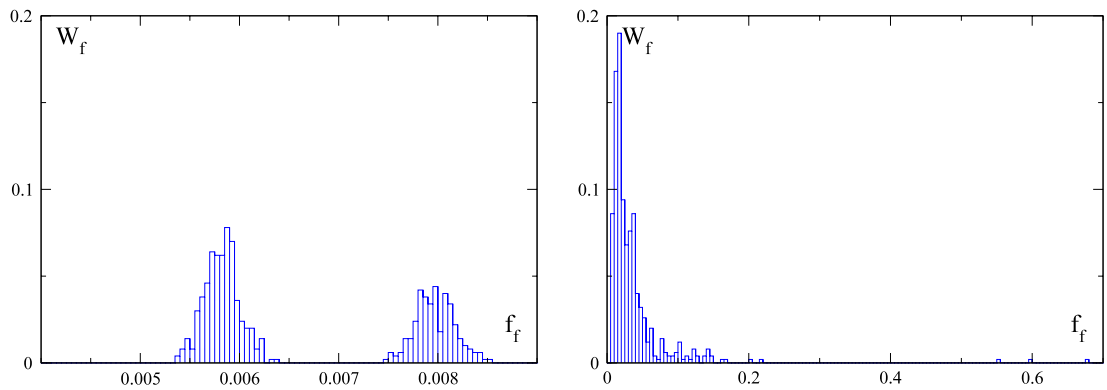
Fig. 15. (Color online) Data for the LiveJournal network: probability distribution W_f of final opinion f_f for a fixed initial opinion $f_i = 0.4$ and $a = 0.5$ in the PROF model. *Left panel:* usual linear weight $P(K)$ in Eq. (1). *Right panel:* a quadratic weight $P^2(K)$ in Eq. (1). Histograms are obtained with $N_r = 500$ initial random realizations; the normalization is fixed by the condition that the sum of W_f over all histogram bins is equal to unity.

own opinion, being different from the main opinion of the majority. In this way, the proposed PROF–Sznajd model shows that totalitarian opinions can be escaped from by small subcommunities. We find that the properties of opinion formation are rather similar for the two university networks of Cambridge and Oxford. However, the results obtained for networks of LiveJournal and Twitter show that the range of bistability practically disappears for these networks. Our data indicate that this is related to a slower algebraic decay of PageRank in these cases compared to the university networks. However, the deep reasons for such a difference require a more detailed analysis. Indeed, the LiveJournal and Twitter networks demonstrate rather different behavior for the P^2-weighted function of opinion formation. The studies performed for regular networks [10] show the existence of stable or bistable fixed points for opinion formation models that have certain similarities with the opinion formation properties found in our studies. At the same time the results obtained in Ref. [38] show that three-body spin coupling can generate a chaotic renormalization dynamics. Some of our results (Fig. 15, right panel) give indications of the possible existence of such a chaotic phase in social networks.

The enormous development of social networks in the last few years [2–5] definitely shows that the analysis of opinion formation on such networks requires further investigations. This research can find also various other applications. One of them could be a neuronal network of a brain which represents itself as a directed scale-free network [40]. The applications of network science to brain networks is now under rapid development (see, e.g., Ref. [41]), and Google matrix methods can find useful applications in this field [42].

Fig. 16. (Color online) Data for the Twitter network. *Left panel:* PageRank probability decay with PageRank index K (full curve); the fitted algebraic dependence is shown by the dashed line $y = b - \beta x$ (for $1 \leq \log_{10} K \leq 5.5$) with the exponent $\beta = 0.511 \pm 0.0021$ and $b = -3.33 \pm 0.0069$ (for the range $5.5 \leq \log_{10} K \leq 7$ we find $\beta = 1.23$). *Right panel:* dependence of the final fraction of red nodes f_f on the tenacious parameter a (or conformist parameter $b = 1 - a$) in the PROF model for initial red nodes in N_{top} values of PageRank index ($1 \leq K \leq N_{top}$; cf. Fig. 4, Fig. 14). Here, $N_{top} = 1200$ (blue line at $f_f = 0$); 1250 (red curve with circles); and 1300 (top green line); $T = 0$.

Acknowledgments

References

[1] J.R. Zaller, The Nature and Origins of Mass Opinion, Cambridge University Press, Cambridge, UK, 1999.
[2] Wikipedia, LiveJournal, March 9, 2012. http://en.wikipedia.org/wiki/LiveJournal.
[3] Wikipedia, Facebook, March 9, 2012. http://en.wikipedia.org/wiki/Facebook.
[4] Wikipedia, Twitter, March 9, 2012. http://en.wikipedia.org/wiki/Twitter.
[5] Wikipedia, VK (Social network), March 9, 2012. http://en.wikipedia.org/wiki/VK_(social_network).
[6] S.N. Dorogovtsev, J.F.F. Mendes, Evolution of Networks, Oxford Univ. Press, 2003.
[7] G. Caldarelli, Scale-Free Networks, Oxford Univ. Press, 2007.
[8] S. Galam, J. Math. Psych. 30 (1986) 426.
[9] T.M. Liggett, Stochastic Interacting Systems: Contact, Voter and Exclusion Processes, Springer, Berlin, 1999.
[10] S. Galam, Europhys. Lett. 70 (2005) 705.
[11] D.J. Watts, P.S. Dodds, J. Consumer Research 34 (4) (2007) 441.
[12] S. Galam, Int. J. Mod. Phys. C 19 (2008) 409.
[13] C. Castellano, S. Fortunato, V. Loreto, Rev. Mod. Phys. 81 (2009) 591.
[14] P.L. Krapivsky, S. Redner, E. Ben-Naim, A Kinetic View of Statistical Physics, Cambridge University Press, Cambridge, UK, 2010.
[15] B. Schmittmann, A. Mukhopadhyay, Phys. Rev. E 82 (2010) 066104.
[16] Academic Web Link Database Project. http://cybermetrics.wlv.ac.uk/database/.
[17] M. Kurucz, A.A. Benczúr, A. Pereszlenyi, Large-scale principal component analysis on livejournal friends network, in: Proc. Workshop on Social Network Mining and Analysis Held in Conjunction with 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, Las Vegas NV, August 24–27, 2008. http://dms.sztaki.hu/en/letoltes/livejournal-data.
[18] H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a social network or a news media? in: Proc. 19th Int. Conf. WWW2010, ACM, New York, NY, 2010, p. 591. the web data are downloaded from the web site maintained by S. Vigna. http://law.dsi.unimi.it/webdata/twitter-2010/.
[19] S. Brin, L. Page, Comput. Netw. ISDN Syst. 30 (1998) 107.
[20] A.M. Langville, C.D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press, Princeton, 2006.
[21] S. Redner, Phys. Today 58 (6) (2005) 49.
[22] F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Phys. Rev. E 80 (2009) 056103.
[23] J.D. West, T.C. Bergstrom, C.T. Bergstrom, Coll. Res. Lib. 71 (2010) 236. http://www.eigenfactor.org/.
[24] F. Radicchi, PLoS ONE 6 (2011) e17249.
[25] A.O. Zhirov, O.V. Zhirov, D.L. Shepelyansky, Eur. Phys. J. B 77 (2010) 523.
[26] L. Ermann, D.L. Shepelyansky, Acta Phys. Pol. A 120 (6A) (2011) A158.
[27] T. Preis, D. Reith, H.E. Stanley, Phil. Trans. R. Soc. A 368 (2010) 5707.
[28] T. Preis, H.S. Moat, H.E. Stanley, S.R. Bishop, Sci. Reports 2 (2012) 350.
[29] S. Saavedra, J. Duch, B. Uzzi, PLoS ONE 6 (2011) e26705.
[30] J. Ginsberg, M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, L. Brilliant, Nature 457 (2009) 1012.
[31] A. Vespignani, Science 325 (2009) 425.
[32] K. Sznajd-Weron, J. Sznajd, Int. J. Mod. Phys. C 11 (2000) 1157.
[33] K.M. Frahm, B. Georgeot, D.L. Shepelyansky, J. Phys, A: Math. Theor. 44 (2011) 465101.
[34] L. Ermann, A.D. Chepelianskii, D.L. Shepelyansky, J. Phys. A: Math. Theor. 45 (2012) 275101.
[35] S. Galam, B. Walliser, Physica A 389 (2010) 481.
[36] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.

[37] V. Sood, S. Redner, Phys. Rev. Lett. 94 (2005) 178701.
[38] N.S. Ananikian, S.K. Dallakian, Physica D 107 (1997) 75.
[39] G.K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley, Boston, 1949.
[40] V.M. Eguiluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, A.V. Apkarian, Phys. Rev. Lett. 94 (2005) 018102.
[41] X.-N. Zuo, R. Ehmke, M. Mennes, D. Imperati, F.X. Castellanos, O. Sporns, M.P. Milham, Cereb. Cortex, 2011. http://dx.doi.org/10.1093/cercor/bhr269.
[42] D.L. Shepelyansky, O.V. Zhirov, Phys. Lett. A 374 (2010) 3206.