Physica A 391 (2012) 5779–5793

Contents lists available at SciVerse ScienceDirect

Physica A

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

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

Article history:
Received 22 April 2012
Received in revised form 18 June 2012
Available online 23 June 2012

Keywords:
Voting
PageRank
Opinion formation

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.

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).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2012.06.047

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 $a + b = 1$ ($0 \le a, b \le 1$). 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 $J_{ij} \ne J_{ji}$, 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