Reversible bootstrap percolation: Fake news and fact checking

Matías A. Di Muro, Sergey V. Buldyrev, and Lidia A. Braunstein

Reversible bootstrap percolation: fake news and fact-checking

Matías A. Di Muro

Instituto de Investigaciones Físicas de Mar del Plata (IFIMAR)-Departamento de Física,

Facultad de Ciencias Exactas y Naturales,

Universidad Nacional de Mar del Plata-CONICET,

*Funes 3350, (7600) Mar del Plata, Argentina.**

Sergey V. Buldyrev

Department of Physics, Yeshiva University,

500 West 185th Street, New York, New York 10033, USA and

Politecnico di Milano, Department of Management,

Economics and Industrial Engineering,

Via Lambruschini 4, BLD 26, 20156 Milano, Italy

Lidia A. Braunstein

Instituto de Investigaciones Físicas de Mar del Plata (IFIMAR)-Departamento de Física,

Facultad de Ciencias Exactas y Naturales,

Universidad Nacional de Mar del Plata-CONICET,

Funes 3350, (7600) Mar del Plata, Argentina. and*

Physics Department, Boston University,

Boston, Massachusetts 02215, USA

Abstract

Bootstrap percolation has been used to describe opinion formation in the society and other social and natural phenomena. The formal equation of the bootstrap percolation may have more than one solution, corresponding to several stable fixed points of the corresponding iteration process. We construct the reversible bootstrap percolation process, which converges to these extra solutions displaying a hysteresis typical of discontinuous phase transitions. This process provides a reasonable model for the fake news spreading and the effectiveness of fact-checking. We show that sometimes it is not sufficient to discard all the sources of fake news in order to reverse the belief of the population which has been formed under the influence of these sources.

* mdimuro@mdp.edu.ar

The spread of information within a population is an interesting phenomena from which we can learn, how prone is a massive group of people to embrace and propagate fake news or conspiracy theories for instance. One of the models that can describe this process is bootstrap percolation, which is a fairly simple threshold model that has been widely studied in the last years to mimic different spreading processes on complex systems. In this model a random fraction of nodes or sites activate or adopt a new idea spontaneously. Then, other nodes subsequently activate if they are connected with at least a minimal number of active neighbors [1]. The initial random activation triggers a cascading process which stops at when the system stabilizes.

This model was first introduced to understand the mechanisms of ferromagnetism on Bethe lattices [2], and then in the following years, it was studied on a variety of graphs [3–5]. Bootstrap percolation, along with other thresholds models such as k-core percolation, are useful to describe social processes, in which people tend to change their opinion if they are influenced by their multiple contacts [6, 7]. Accordingly, these models can potentially describe phenomena such as the spreading of gossip or fake news, viral marketing and opinion formation [8–10]. Also, people tend to adopt new technologies or brands when they are in contact with people that are already using them [11, 12]. Bootstrap percolation has also non-social applications, such as the study of fault tolerance in distributed computing [13] and cascading failures in power grids or communication networks. Furthermore, the spreading of a disease and the diffusion of awareness [14] can be studied using these kind of threshold models [15].

In bootstrap percolation, a fraction f of nodes are spontaneously activated at the initial stage of the process. Such nodes are called "seeds", while the rest of them are called "non-seeds". A non-seed node with degree k needs to be supported or influenced by at least $k^* \leq k$ active neighbors to be activated. The values of k and k^* may be different for different nodes and we will assume that they are randomly chosen from the degree distribution $P(k)$ and a threshold distribution $r(j, k)$, respectively, where $r(j, k)$ is the cumulative distribution function of the threshold which is the probability of finding a node with $k^* \leq j$, given that it has degree k. The activation of the seed nodes leads to a cascade of activation at the end of which the fraction $S \geq f$ of nodes become active. This fraction can be regarded as an order parameter of the model, and at certain $f = f_t$ may undergo a discontinuous transitions similar to crossing a spinodal associated with the first order phase transitions

in condensed matter, jumping from a small value for $f < f_t$ to a larger value for $f > f_t$. For example, in the liquid-gas phase transition there are two spinodals (lines of diverging compressibility) emanating from the critical point[16]. When approaching spinodals, the uniform metastable phases (supercooled gas or superheated liquid) reach their stability limits and immediately phase segregate into a mixture of two phases.forming droplets of liquid or bubbles of gas. The authors in network science called such a transition a hybrid transition [4]because approaching this transition from one side the derivative of dS/df diverges and the removal of a single node causes a power law distributed avalanches, thus having features of both first and second order phase transitions. However the same phenomenology is present near the spinodal: isothermal compressibility $(-\partial V/\partial P)_T/V$ diverges, and the uniform density phase has diverging density fluctuations, before it breaks down and phase segregate. The only difference is that in networks, the actual line of equilibrium first order phase transition defined by the condition of equal chemical potentials of the two phases is not observable, since the thermodynamic potential is not defined.

The bootstrap percolation model can be solved exactly in the limit of infinitely large networks randomly connected with a given degree distribution [4, 17], when the probability of short loops is negligible. At the end of activation cascade for given f the fraction of active nodes can be written as

$$S = f + (1 - f)\Psi(Z),\tag{1}$$

where $\Psi(Z)$ is the bootstrap generating function

$$\Psi(Z) = \sum_k P(k) \sum_{j=0}^{k} r(j, k) \binom{k}{j} Z^j (1 - Z)^{k-j}.\tag{2}$$

and probability Z satisfies a self-consistent equation

$$Z = f + (1 - f)\Phi(Z),\tag{3}$$

where

$$\Phi(Z) = \sum_k \frac{kP(k)}{\langle k \rangle} \sum_{j=0}^{k-1} r(j, k) \binom{k-1}{j} Z^j (1 - Z)^{k-j-1},\tag{4}$$

is the bootstrap generating function for the excess degree distribution, $\langle k \rangle$ is the average degree of the network and Z is the probability of reaching via a random link a seed node or a non-seed node supported by at least k^* of its $k - 1$ outgoing neighbors. The nonlinear

4

equation Eq. (3), may have more than one solution which can be obtained by an iteration method corresponding to the stages of activation/deactivation process:

$$Z_{n+1} = f + (1 - f)\Phi(Z_n), \tag{5}$$

where $Z_n \to Z$ for $n \to \infty$. For a fixed fraction of seed nodes f, if the initial value of $Z = Z_0$ of Eq. (5) is small enough, e.g. $Z_0 \leq f$, then the iterations converge to the smallest stable fixed point Z_I corresponding to the direct bootstrap percolation. On the other hand, if the initial value of Z is large enough e.g. $Z_0 > 1 - \epsilon$, where $\epsilon > 0$, the iterations converge to a different stable fixed point $Z_{II} \geq Z_I$ if such a point exists. For many reasonable degree and threshold distributions, function $y = f + (1 - f)\Phi(x)$, has inflection points and crosses the straight line $y = x$ several times, producing several stable fixed points of the iterative process (5), when at the crossing point $(1 - f)d\Phi(x)/dx < 1$. In the limit of infinitely large network, if $Z_0 = 0$, and the fraction of active nodes S_0 coincides with the fraction of seeds $S_0 = f$, equation

$$S_n = f + (1 - f)\Psi(Z_n) \tag{6}$$

gives the fraction of active nodes after n-th stage of the activation cascade such that at each stage each inactive node i counts its already active neighbors and if this number $k_i^a \geq k_i^*$, node i activates. However, it is not clear, how to solve this problem analytically for other initial condition of the network. It is tempting to suggest that the second fixed point corresponds to a reverse process associated with a hysteresis phenomenon typical to discontinuous phase transitions in general and network science in particular [18, 19]. One potential candidate could be k-core percolation [20, 21]. In this process, all nodes are initially active when a random fraction $1 - p$ of them fails spontaneously . Each node has a functionality threshold, i.e., a minimum number of neighbors k_c^* that must be active to keep it in that state. Thus, the random failure generates a pruning process that may end up affecting a significant amount of nodes in the network. Even though bootstrap and k-core are quite similar, it has been shown that they are not opposite processes but complementary under the right conditions [17]. In particular, the k-core process with $p = 1 - f$ and complimentary threshold distribution is not a reverse process for the bootstrap percolation but describes the decrease of the number of inactive nodes in the same bootstrap percolation process. In general, it is impossible to construct a k-core process such that it will describe the reverse bootstrap process.

The problem with Eqs. (1-5) is that while the physical meaning of S is clear: it is the fraction of active nodes in the system and it is easy to construct a cascading process of consecutive activation of nodes in which the initial set of active nodes A_0 coincides with the seed nodes and the set of active nodes A_n at the n-th stage of the cascade can be readily determined by the set A_{n-1}, the physical meaning of the probability Z_n is less clear. In Ref. [17] it was defined as the probability of reaching an already activated node or a seed by a randomly selected directed link, emanating from any node, active, inactive, or seed. But this definition does not link Z_n to any material feature of the network and does not specify, how it can be computed for a given network configuration.

The aim of this paper is to find the physical meaning of Z_n and construct the inverse process for bootstrap percolation that converges to the larger fixed point of Eq. (5). In addition, in the case when Eq. 3 has a unique solution Z_s, the process must reach that solution starting from an initial condition Z_0 such that $Z_0 > Z_s$. To achieve this goal, we develop a reversible bootstrap percolation model, which dynamically responds to the changes in the seed configuration, leading to activation of nodes if the fraction of seeds increases and to deactivation of nodes if the number of seeds decreases. We assume that during this process some nodes can spontaneously become active without being surrounded with at least k^* neighbors, thus becoming seeds which are self-sufficient, while some nodes can spontaneously lose this property, i.e. become regular nodes without being immediately deactivated. From the social perspective, these nodes change from agents actively influencing their neighbors by spreading news or advertising new products (for example, because they are paid for doing this) to regular members of a community whose beliefs and habits are influenced solely by the beliefs and habits of their contacts in the social network. The reduction of the number of seed nodes and the subsequent deactivation of other nodes can be associated with the process of fact checking.

We must state that our current model may not be adequate for real social phenomena, because we deliberately assume that the network topology does not change during this process. In reality, it may not be true. The agents involved in these activities can be simply deleted from the network or inserted into it. We ignore such events, to keep the model analytically tractable.

In our model of reversible bootstrap percolation, at any moment of time t, we define a network of directed *influence* links, going along the static links of the network (Fig. 1). A

static link may correspond to (a) two influence links going in the opposite directions, (b) to a single influence link going in a single direction, or (c) to no influence links at all. We define Z_t as the number of influence links $N_i(t)$ divided by the doubled number of static links: $Z_t = N_i(t)/(2N_l)$. The status of any non-seed node, i at any moment of time is defined by the number of influence links, $k'(t)$ it receives: if this number is greater or equal than the activation threshold of this node, $k_i'(t) > k_i^*$: the node is active, otherwise it is inactive. The seed nodes are always active, however as any other nodes, they also receive influence links, and this number starts to determine their status as soon as they spontaneously lose their seed property. If, in contrast a non-seed node becomes seed, it immediately activates, its k_i does not change, but it stops to determine its status.

The initial status of the influence network and seed nodes can be arbitrary, but for clarity we assume that at $t = 0$ they are empty: $f(0) = Z(0) = 0$. Then some seed nodes spontaneously appear, and the dynamic process of activation and self-pruning of influence network begins (Fig 1). From a social perspective, creation of an influence link corresponds to news spreading, while its deletion corresponds to fact checking. In principle, the process of deletion and creation can be random: at any moment of time, an agent represented by a node starts to examine its influence links. It selects a static link, determines a neighbor connected by this link and examine the strength of this neighbor belief. If this neighbor is a seed, the influence link leading from the seed to this agent is created (if it has not existed before). If the neighbor is a non-seed, the agent counts the number of influence links, $k_n'(t)$ the neighbor receives excluding the influence link, coming from the agent. This exclusion is crucial for the fact-checking, the agent wants to be sure that the neighbor has a firm belief without an influence from the agent. This exclusion corresponds to the word "already" in the definition of Z in Ref. [17]. In Eq. (4) the exclusion is reflected by the reduction of the index of summation from k to $k - 1$ in the inner sum over j: here $k - 1$ is the number of outgoing links of the neighbor, excluding one link by which this neighbor was reached from the agent. If $k_n' \geq k_n^*$, the agent creates an influence link from this neighbor, if it has not existed before. If in contrast, $k_n' < k_n^*$, the agent deletes an influence link from this neighbor, if it has existed. By this method, which represents, both news spreading and fact checking, the influence network is kept updated. It is obvious that this method exactly corresponds to the iterative Eq. (5), if at each time step, n, the status of all influence links is checked simultaneously. Computationally, we keep two arrays of influence links I_{n-1} and

7

I_n, array I_n at first is empty and is created using existing array I_{n-1}. Once I_n is created, the fraction of influence links Z_n is computed and the set, A_n, of active nodes is determined as the nodes for which the number of influence links they receive, $k_i^i \geq k_i^*$. This determination corresponds to the computation of the fraction of the active nodes S_n using Eq. (6). Note, that in the direct bootstrap process, described above, $k_i^a = k_i^i$, and there is no need to introduce the influence network to describe the activation process of nodes starting from $Z_0 = 0, S_0 = f$. However, once the fraction of seed nodes change, especially if it decreases, the network of influence links must be used to construct a reversible bootstrap process, in which the fraction of active nodes reduces in response to the reduction of the number of seeds. We see that the dynamic process of the influence link updates together with the set of seeds completely determines the set of active nodes at each stage. Conversely, the influence network can be constructed for any initial set of active nodes A_0 and a set of seed nodes, F_0, by the same update process described above if we begin our iterations with an influence network consisting of directed links emanating from each active node. During this process which resembles pruning, certain influence links are removed, as a result, certain nodes are deactivated and the influence network converges to a steady state, described by Eqs. (3) and (1).

If the seed nodes change infrequently comparatively to the influence network update, that is, after each change Z_n stabilizes to one of the stable fixed points of Eq. (3), then at each stage n the values of Z_n and S_n predicted by equations (5) and (6) coincide with the simulation results for a large enough network.

If the seeds change frequently before Z_n stabilizes, we have a dynamical process similar to one presented in [19], with the difference that reversible bootstrap satisfies exact equations for Φ and Ψ, while in [19] the equations are approximate in the limit of large $\langle k \rangle$, because in this limit function Ψ, used in Ref. [19], approximates function Φ. The dynamic process introduced here can be applied for many social and natural phenomena and has rich possibilities for generalization such as varying the speed of change of seed nodes, changing the method of updating influence network, and even a possibility of changing static links and activation thresholds. Also it is obvious that one can construct a reversible k-core process using a reversible complementary bootstrap process.
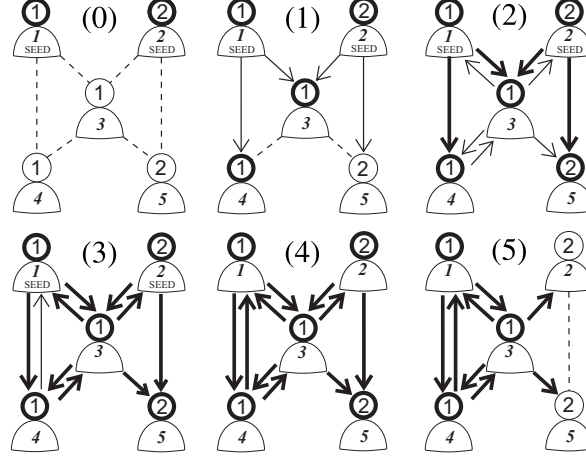
FIG. 1: Dynamic updates during reversible bootstrap percolation. Stage (0): two seed nodes are activated. The static links are shown by dashed lines. Each node has its activation threshold on its head, and its identification number on its body. Active nodes shown by bold circles around their heads. Seed nodes are active by definition. $Z_0 = 0$, $S_0 = 2/5$. Stage (1): the influence links from the seeds are created, forming set I_1. $Z_1 = 4/12$. Nodes 3 and 4 become active because the number of influence links they have received is greater than or equal to their thresholds. $S_1 = 4/5$. Stage (2): Old influence links (I_1) are shown by bold arrows. New influence links (I_2) obtained by the rules of the model from I_1 are shown by thin arrows. Node 3 receives influence from node 4, because node 4 does not depend on node 3 to stay at the threshold. However, seed 1 does not receive the influence from node 4 because the latter does depend on the former to stay at the treshold. In contrast, node 3 influence all of its neighbors, because from the point of view of any of its neighbors, node 3 is active without its influence. Node 5 now receives sufficient number of influence links to become activated. $Z_2 = 9/12$, $S_2 = 1$. Stage (3): Seed 1 receives influence from node 4, because now node 4 has additional influence from node 3 and stays at the threshold without the influence of seed 1. Node 5 can not send influence links to any of its neighbors, because it would be inactive without the support of each particular neighbor. $Z_3 = 10/12$, $S_3 = 1$. Stage (4): The influence network comes to a steady state, but both seeds spontaneously lose their seed properties. $Z_4 = 10/12$, $S_4 = 1$. Stage (5) Node 2 is deactivated, because its number of influence links is below the threshold. The influence links that come from node 2 are removed. Node 1 stays at the threshold active and influence all of its neighbors, because it stays at the threshold without the support of each of them. Node 5 falls below the threshold and is deactivated. The network comes to a new steady state. $Z_5 = 8/12$, $S_5 = 3/5$.

The theoretical prediction of the reversible bootstrap percolation model is in excellent agreement with simulations of finite graphs (Fig. 2). Here we have created a sequence of seed-changing events f_t separated by large enough time intervals so that between them the network comes to a steady state. That is, once the number of seeds change it remains the same until the fake news spreading or the fact checking process finishes. We see that the network strictly follows the hysteresis loop predicted by the theory. In these simulations we use a network with the Poisson degree distribution, and a rather complex threshold distribution $r(j,k) = F_\gamma(j/k)$ with where function $F_\gamma(x) = [x^2(18\gamma^2 - 12\gamma) + x^3(4 - 12\gamma^2) + x^4(6\gamma - 3)]/(6\gamma^2 - 6\gamma + 1)$ has an inflection point at $x = \gamma$. The reason of using such a complex function is two-fold. First we want to demonstrate that the model works for an arbitrary threshold distribution. Indeed, if the distribution of the activation thresholds can be obtained empirically for certain social or economic networks, the model can use these empirical distributions. But for this particular $r(j,k)$, which is a polinomial of j/k, the functions Φ and Ψ can be computed analytically (see Supplementary information for Ref.[22]). Second, we want to demonstrate that a systematic change of the inflection point of the threshold distribution dramatically changes the outcome of the model which has a non-trivial phase diagram in the (f, γ, S) space (Fig. 3), which corresponds to the phase diagram of the Van der Waals model (see eg. [16]) in the (P, T, ρ) space with a critical point, two spinodals and the mean-field critical exponents: $\beta = 1/2$, $\gamma = 1$, $\delta = 3$. In this case, the ordering field, pressure P, corresponds to the fraction of seeds, f, which is (using political terms) is the intensity of fake news creation. The thermal field, temperature T, corresponds to γ, which models the resistance of the population to the fake news. Lower inflection point in the $r(j,k)$ distribution indicates lower threshold to the fake news. Finally the density ρ is analogous to S because both are order parameters. Similar phase diagram appears if we keep γ constant but vary $\langle k \rangle$.
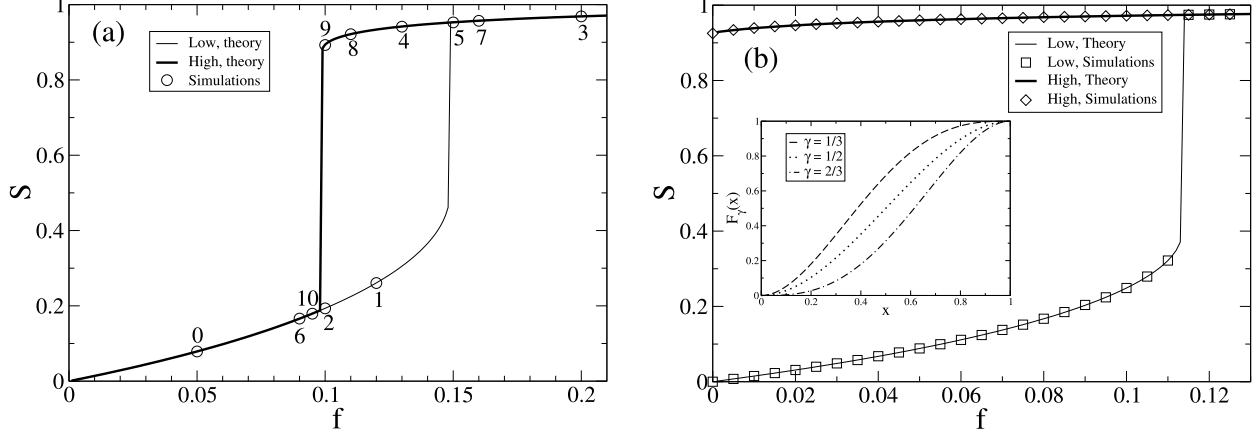
FIG. 2: Comparison of the theory and simulations for the reversible bootstrap simulation model. In both panels the degree distribution $P(k)$ is Poisson with $\langle k \rangle = 10$ and $r(j,k) = F_\gamma(j/k)$ with $\gamma = 0.53$ (a) and $\gamma = 0.48$ (b). In the inset we show the function $F_\gamma(x)$ for $\gamma = 1/3$, $\gamma = 1/2$ and $\gamma = 2/3$. In both cases the size of the networks is $N = 10^7$ nodes. In (a) the simulations follow a reversible bootstrap process based on the influence network updates described above for a sequence of f_t, marked on the graph by numbers $t = 0, 1, 2, ..., 10$. For (b), the simulations are done for increasing f from $f = 0$ to $f = 0.125$ giving the low solution and then in reverse order from $f = 0.125$ to $f = 0$ giving the high solution. One can see that in panel (b) the fraction of active nodes once it undergoes transition to a high activated phase at $f = f_t = 0.114$ never returns to a low activated phase even for $f = 0$. One can see that a small increase in susceptibility of the population to the fake news modeled by a small decrease of the inflection point in the distribution $r(j,k)$, may lead to the irreversible opinion change of the population. We obtained similar results for many other types of distributions $P(k)$ and $r(j,k)$.
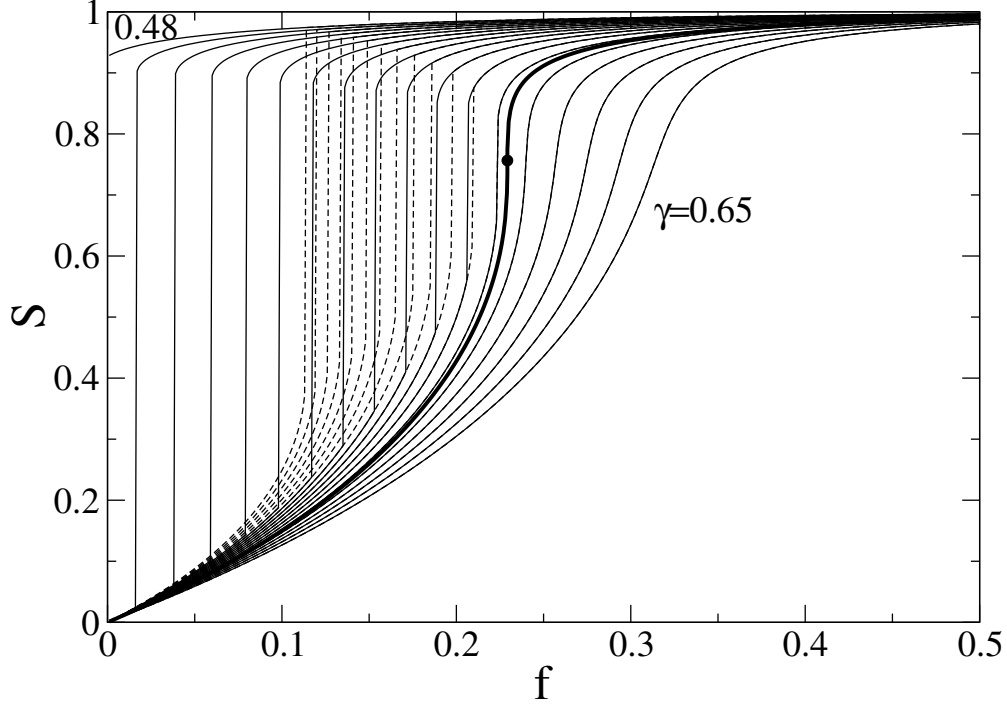
FIG. 3: Phase diagram of the reversible bootstrap model, in the (f, S) plane which is completely analogous the Van der Waals phase diagram in the (P, ρ) plane. The inflection point of the threshold distribution is analogous to temperature. The black line corresponds to isochores (lines of equal γ with step .01 from $\gamma = 0.48$ to $\gamma = 0.65$) following the high stable point of Eq. (5) physically analogous to the superheated metastable liquid crossing the liquid gas spinodal. The dashed lines follow the low stable solution, physically corresponding to the metastable supercooled gas phase crossing gas liquid spinodal. Above the critical point $\gamma = 0.6035$, $f = 0.2292$, $S = 0.754$, the second stable fixed point of Eq. (5) disappears.

A nontrivial outcome of this model is that once the public opinion is switched by spreading fake news to a new state undergoing a spinodal crossing, it may not return to the original state even when all the fake news sources are discarded. Moreover, for certain distributions $P(k)$ and $r(j, k)$, [Fig. 2 (b)] the fraction of active nodes will not undergo a reverse phase transition and will not return to a low-fraction state, even if all seeds are eliminated as demonstrated also in a schematic Fig. 1 for a very small system. In Fig. 2 (b) the degree distribution is the same, but the bootstrap thresholds are slightly lower than in Fig. 2 (a). In political language, the fake news now are closer to the hearts of the population than in Fig. 2 (a), so the population accepts them easier and keeps believing them even after all of

them have been firmly discarded and all sources of them are deactivated.

In conclusion, we construct and investigate a model of reversible bootstrap percolation which can be applied to political and social science, especially to the problem of fake news spreading and efficiency of the fact checking. This model shows that misconceptions that has been established in a population may prevail, even when many of the primary sources that disseminated them are debunked.

---

[1] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99** (2002).

[2] J. Chalupa, P. L. Leath, and G. R. Reich, J. Phys. C: Solid State Phys. **12**, L31 (1979).

[3] S. Janson *et al.*, Electron. J. Probab. **14**, 86 (2009).

[4] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **82**, 011103 (2010).

[5] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **83**, 051134 (2011).

[6] D. Centola, Science **329**, 1194 (2010).

[7] M. O. Jackson and D. López-Pintado, Netw. Sci. **1**, 49 (2013).

[8] D. Kempe, J. Kleinberg, and É. Tardos, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2003) pp. 137–146.

[9] P. Domingos and M. Richardson, in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2001) pp. 57–66.

[10] M. Shrestha and C. Moore, Phys. Rev. E **89**, 022805 (2014).

[11] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).

[12] D. Kempe, J. Kleinberg, and É. Tardos, in *Automata, Languages and Programming* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005) pp. 1127–1138.

[13] S. Kirkpatrick, W. W. Winfried, G. B. Robert, and H. Harald, Physica A **314** (2002).

[14] C. Granell, S. Gómez, and A. Arenas, Phys. Rev. E **90**, 012808 (2014).

[15] L. Feng, Y. Hu, B. Li, H. E. Stanley, S. Havlin, and L. A. Braunstein, PLOS ONE **10**, 1 (2015).

[16] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. E **83**, 065101(R) (2011).

[17] M. A. Di Muro, L. D. Valdez, H. E. Stanley, S. V. Buldyrev, and L. A. Braunstein, Phys. Rev. E **99**, 022311 (2019).

[18] B. Min and K. I. Goh, Phys. Rev. E **89**, 040802(R) (2014).

[19] A. Majdandzic, B. Podobnik, S. V. Buldyrev, D. Y. Kenett, S. Havlin, and H. E. Stanley, Nature Physics **10**, 34 (2014).

[20] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. Lett. **96**, 040601 (2006).

[21] N. Azimi-Tafreshi, S. Osat, and S. N. Dorogovtsev, Physical Review E **99**, 022312 (2019).

[22] M. A. Di Muro, L. D. Valdez, H. H. Aragão Rêgo, S. V. Buldyrev, H. E. Stanley, and L. A. Braunstein, Sci. Rep. **7**, 15059 (2017).