Physica A 388 (2009) 3901–3910

Contents lists available at ScienceDirect

Physica A

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

A modified Galam's model for word-of-mouth information exchange

Andrea Ellero *, Giovanni Fasano, Annamaria Sorato

Università Ca' Foscari di Venezia, Dipartimento di Matematica Applicata, Ca' Dolfin - Dorsoduro 3825/E - 30123, Venezia, Italy

ARTICLE INFO

Article history:
Received 1 October 2008
Received in revised form 4 May 2009
Available online 6 June 2009

Keywords:
Word-of-mouth process
Information spreading

ABSTRACT

In this paper we analyze the stochastic model proposed by Galam in [S. Galam, Modelling rumors: The no plane Pentagon French hoax case, Physica A 320 (2003), 571–580], for information spreading in a 'word-of-mouth' process among agents, based on a majority rule. Using the communications rules among agents defined in the above reference, we first perform simulations of the 'word-of-mouth' process and compare the results with the theoretical values predicted by Galam's model. Some dissimilarities arise in particular when a small number of agents is considered. We find motivations for these dissimilarities and suggest some enhancements by introducing a new parameter dependent model. We propose a modified Galam's scheme which is asymptotically coincident with the original model in the above reference. Furthermore, for relatively small values of the parameter, we provide a numerical experience proving that the modified model often outperforms the original one.

1. Introduction

The dynamics of social contagion, in particular of opinion spreading on a population, has been studied in many contexts, with many different approaches and diverse applications in social sciences, experimental psychology, consumer research, finance (see e.g. Refs. [2–5] and the references therein). The significance of word-of-mouth information and rumor diffusion through social networks has been widely recognized as fundamental in all these contexts, see e.g. Refs. [6–10].

The leverage effect of word-of-mouth information on consumer behavior, for example, improves the effectiveness of communication activities of a firm [11]. Word-of-mouth effects cannot be ignored as a powerful marketing tool, especially in the case of new product introductions, when the aim is to reduce the probability of a post-launch failure [12].

We focus on a diffusion model which investigates opinion dynamics, driven by rumors in a population, proposed by Galam [1,13,14]. He considers a population of agents who can have two opposite opinions, for example about political elections. Each of them can shift their opinion to the opposite one, due to repeated discussions within a group of people, following a majority rule which is biased in favor of one of the two opinions in the case of parity. The model provides an explanation of the spreading of the rumor claiming that "No plane did crash on the Pentagon on September 11". Interestingly, for this agent based model Galam provides a closed form formula for the probability of one of the opinions to prevail, at a certain time instant.

Galam's model has been extended, for example considering agents having three possible opinions [15], and applied in several contexts, such as fashion industry [16] or political elections [8]. For a comprehensive review see Ref. [17].

In this paper we study the spreading of opinions in the special case when the number of interacting agents is relatively small. This case is remarkable in practice. Consider for example the spreading of an opinion among members of a board of directors, of a municipal administration, of a faculty, or among consumers discussing about niche or highly technological products. In fact "research estimates that group-based work methods exist in nearly 70% of US firms" [18,19]. The size of the group

* Corresponding author.
E-mail addresses: ellero@unive.it (A. Ellero), fasano@unive.it (G. Fasano), amsorato@unive.it (A. Sorato).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2009.06.002

affects the dynamics of the opinion spreading: "verbal brainstorming groups, for example, should experience synergy that increases as group size increases, and some social facilitation effects, but also suffer from process losses that increase with the size of the group due to production blocking, social loafing, evaluation apprehension, and cognitive interference" [20].

We first simulated the rumor spreading following Galam's guidelines [1]. Observing the outcomes, we figured out that, considering a relatively small number of agents, Galam's model may possibly return inaccurate results. Then, in order to cope with the latter drawback, we proposed a modified Galam's scheme, which is asymptotically coincident with the original model. When few agents are considered the modified model often turns out to be preferable to the original one; i.e., a more accurate estimation of the simulated data is obtained.

This paper is organized as follows: Section 2 introduces Galam's model, details about its peculiarities and some possible limits are provided in Section 3. Then, in Section 4 we describe our model, including a motivated numerical comparison with Galam's scheme. Finally, Section 5 completes the paper with suggestions for future work.

2. Galam's model

Galam [1] considers a population whose individuals can change opinion after discussing in groups. Let N be the overall number of people who meet into groups, in order to exchange their information. Each people either thinks, say, '$+$' or '$-$' and $N = N_+(t) + N_-(t)$, where $N_+(t)[N_-(t)]$ is the number of people who respectively think '$+$' ['$-$'] at time step t.

At time step t the N people gather into k-sized groups, $k = 1, \dots, L$, with probability a_k $(a_1 + \cdots + a_L = 1)$. Then, after a discussion in each group, at time $t + 1$ people can shift their opinion to the opposite one, following a *majority rule*. The rule of shifting opinion is biased in favor of the opinion '$-$', in case of parity.

Let $P_+(0)$ be the probability to find people thinking '$+$' at time step '0', and let

$$P_+(0) = \frac{N_+(0)}{N}; \qquad P_-(0) = 1 - P_+(0).$$

Then, Galam's model is described by the following formula

$$P_+(t+1) = \sum_{k=1}^{L} a_k \sum_{j=\left\lfloor \frac{k}{2}+1 \right\rfloor}^{k} C_j^k P_+(t)^j \{1 - P_+(t)\}^{k-j}, \tag{2.1}$$

where $\lfloor x \rfloor$ indicates the largest integer which approximates x from below, and C_j^k is the binomial coefficient. Observing (2.1), we remark that the quantity $P_+(t + 1)$ does not depend explicitly on the number N of interacting people. As in Ref. [1], the dynamic expression (2.1) may be drawn in the space $(P_+(t), P_+(t + 1))$, and can be used to compute a fixed point, which approximates a so called, *killing point*. The killing point can be defined as the theoretical threshold $\bar{P}_+$ such that

$$\begin{aligned} &\text{if } P_+(0) > \bar{P}_+ \text{ then } \lim_{t\to\infty} P_+(t) = 1, \\ &\text{if } P_+(0) = \bar{P}_+ \text{ then } P_+(0) = P_+(t), \text{ for each time step } t > 0, \\ &\text{if } P_+(0) < \bar{P}_+ \text{ then } \lim_{t\to\infty} P_+(t) = 0. \end{aligned}$$

3. Simulation of Galam's model

We have implemented the Galam's process of information exchange by a FORTRAN 90 code (we adopted Compaq Visual Fortran 6.6). The simulation of Galam's model requires a generator of uniformly distributed random numbers, which is usually required to have a long period, low serial correlation, and a tendency not to 'fall mainly on the planes' (see also Ref. [21]). FORTRAN 90 provides the intrinsic routines `RANDOM_NUMBER` and `RANDOM_SEED` for the generation of a uniform distribution of pseudo-random numbers. The routine `RANDOM_NUMBER` implements a modification of a well known linear congruential generator by G. Marsaglia, which is combined with other three random number generators. Routine `RANDOM_NUMBER` is often considered satisfactory (see also Ref. [22]) for general purposes; however, it may be improper when invoked in parallel computing. Moreover, as reported in Ref. [23], `RANDOM_NUMBER` "may contain tendencies in the beginning". In order to avoid the latter drawback, in all the tables of this paper we use `RANDOM_NUMBER` and the first 5000 random numbers generated are disregarded (see Ref. [23]).

To perform the simulation we consider N agents, each of which is an element of a binary string (either '$+$' or '$-$'). According with the initial probability $P_+(0)$, exactly $N_+(0)$ randomly chosen elements of the string are set to '$+$'. We label each entry (agent) of the string, as belonging to a group of size k with probability a_k. In other words, we associate the random number b_i to the i-th entry: then, the i-th entry will belong to a k-sized group if $\sum_{j=1}^{k-1} a_j < b_i \le \sum_{j=1}^{k} a_j$. As a consequence, the string is sub-divided into sub-strings whose length is exactly k, with $k = 1, \dots, L$. Inside each substring, the number of '$+$' is updated according with the majority rule as described by Galam, so that $N_+(1)$ is the overall number of entries '$+$' at iteration 1. Finally, we compute the quantity $N_+(1)/N$. We give an example.

Suppose $N = 10$, $P_+(0) = 0.7$, with $a_2 = a_3 = 0.5$. Now, at $t = 0$, we randomly generate a N-digit string of agents, say '$\{+ - + + + + - + - +\}$', along with the N random numbers $b_1, \dots, b_{10}$. If $b_i \le 0.5$ then we associate the integer '2'

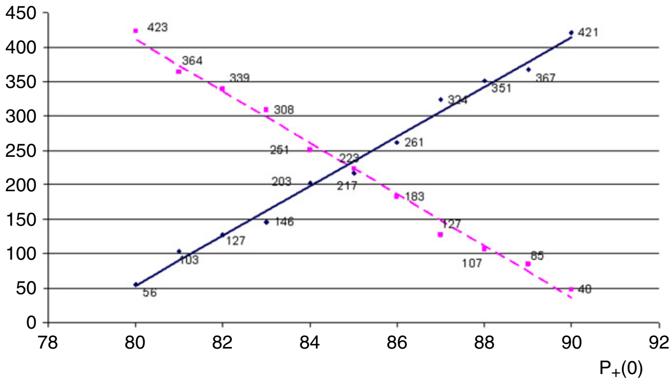

Fig. 1. Results from 500 simulations. For $P_+(0) = 80\%, 81\%, \ldots, 90\%$, two points represent the number of simulations respectively converging to 1 (continuous regression line) and 0 (dashed regression line), within 20 time steps. Following the example in Ref. [1], we set in the simulations $N = 100$, $L = 6$, $a_1 = a_5 = a_6 = 0$ with $a_2 = a_3 = a_4 = 1/3$. The killing point $\bar{P}_+$ is evidently nearby 85%, as suggested in Ref. [1].

Table 1
Average results with $N = 100$, $P_+(0) = 80\%$, 500 runs, Conv-1 $= 500$, Conv-0 $= 0$, $L = 20$, $a_k = 1/20$, $k = 1, \ldots, L$. The column $|\Delta|$_Gal reveals that the model (2.1) may be imprecise.

t-step	Gal	Simul	$\|\Delta\|$_Gal
1	0.8000	0.8000	0.0000
2	0.9392	0.8868	0.0524
3	0.9891	0.9513	0.0379
4	0.9983	0.9819	0.0164
5	0.9997	0.9945	0.0053
6	1.0000	0.9984	0.0015
7	1.0000	0.9995	0.0005
8	1.0000	0.9999	0.0001
9	1.0000	1.0000	0.0000
10	1.0000	1.0000	0.0000

to the i-th agent of the string, otherwise we associate '3'. Let '{2323323222}' be the resulting sequence. Now we collect 2s and 3s into 2-sized and 3-sized groups respectively, so that the string of agents will be correspondingly subdivided into the sub-strings $\{++\}$, $\{-++\}$, $\{++\}$, $\{-+\}$, $\{-00\}$ (00 means that the second 3-sized group has two positions unfilled). Then, after applying the majority rule inside each sub-string (bias for '$-$' in case of ties), we obtain $\{++\}$, $\{+++\}$, $\{++\}$, $\{--\}$, $\{-00\}$, so that at $t = 1$ the new string of agents will be '$\{+++++++-+--\}$'. Observe that one '$+$' and one '$-$' of the initial N-digit string, respectively moved to '$-$' and '$+$'.

If N is finite, the expected number of agents with label k is $N_k = a_k N$: in general N_k might not be a multiple of k.

In order to test Galam's model with a relatively small number of agents (N), in Fig. 1 we provide the results of 500 independent simulations, which reveal the killing point. We will focus on the latter issue in Section 4. The example in Fig. 1 was described in Ref. [1] (parameters of the simulations are detailed in the caption). Repeated simulations (see also Ref. [15]) reveal that the model (2.1) may be inaccurate when N is relatively small: an example of such a behavior is given in Table 1. We remark that results reported in the tables of this paper are averaged over 500 runs (apart from Table 13). Moreover, we set $P_+(0) = 80\%$, following the guidelines in Ref. [1]. Finally, our tables consider the following positions:

- numerical results are performed with the precision of 10^{-16}, though they are reported with just 4 exact digits. In those tables reporting comparisons (Tables 4–13), strictly better results are bolded;
- 'Conv-1' is the number of runs in which the final stationary point is '1', i.e. all the people think eventually '$+$';
- 'Conv-0' is the number of runs in which the final stationary point is '0', i.e. all the people think eventually '$-$';
- 't-step' is the time step (we allowed up to 20 time steps as in Ref. [1]);
- 'Gal' is $P_+(t)$ provided by the Galam model (2.1), at any time step;
- 'Gal$_{\mathcal{M}}$' (Tables 4–13) is $P_+(t)$ provided by the modified model (4.3), at any time step;
- 'simul' is the average ratio $N_+(t)/N$ obtained from the simulations, at time step t;
- '$|\Delta|$-Gal' is given by the quantity | Gal-simul |, and measures the *displacement* between 'Gal' and 'simul';
- '$|\Delta|$-Gal$_{\mathcal{M}}$' (Tables 4–13) is given by the quantity | Gal$_{\mathcal{M}}$-simul |, and measures the *displacement* between 'Gal$_{\mathcal{M}}$' and 'simul'.

Observe from Table 1 that when N is relatively small, the model (2.1) possibly provides inaccurate results with respect to the simulation. We will give a possible explanation of the latter fact, after introducing an improvement of the model in Section 4.

Table 2
Average results with $N = 500$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 472, Conv-0 = 0, $L = 5$, $a_k = 1/5$, $k = 1, \ldots, L$. The model (2.1) performs pretty well.

t-step	Gal	Simul	$\|\Delta\|$_Gal
1	0.8000	0.8000	0.0000
2	0.8195	0.8190	0.0004
3	0.8418	0.8410	0.0008
4	0.8665	0.8647	0.0018
5	0.8921	0.8895	0.0026
6	0.9169	0.9130	0.0039
7	0.9392	0.9334	0.0058
8	0.9576	0.9514	0.0062
9	0.9717	0.9645	0.0072
10	0.9818	0.9752	0.0066
11	0.9885	0.9835	0.0050
12	0.9929	0.9886	0.0043
13	0.9957	0.9927	0.0030
14	0.9974	0.9954	0.0020
15	0.9984	0.9973	0.0012
16	0.9990	0.9984	0.0006
17	0.9994	0.9990	0.0004
18	0.9997	0.9994	0.0003
19	0.9998	0.9996	0.0002
20	0.9999	0.9998	0.0001

Table 3
Average results with $N = 500$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 500, Conv-0 = 0, $L = 10$, $a_k = 1/10$, $k = 1, \ldots, L$. The model (2.1) performs pretty well.

t-step	Gal	Simul	$\|\Delta\|$_Gal
1	0.8000	0.8000	0.0000
2	0.8856	0.8849	0.0008
3	0.9514	0.9490	0.0024
4	0.9833	0.9814	0.0019
5	0.9947	0.9936	0.0011
6	0.9984	0.9979	0.0005
7	0.9995	0.9994	0.0002
8	0.9999	0.9997	0.0001
9	1.0000	0.9999	0.0000
10	1.0000	1.0000	0.0000
11	1.0000	1.0000	0.0000
12	1.0000	1.0000	0.0000
13	1.0000	1.0000	0.0000
14	1.0000	1.0000	0.0000
15	1.0000	1.0000	0.0000
16	1.0000	1.0000	0.0000
17	1.0000	1.0000	0.0000
18	1.0000	1.0000	0.0000
19	1.0000	1.0000	0.0000
20	1.0000	1.0000	0.0000

On the other hand, for L small (say in the range $5 \le L \le 10$) and N relatively large, we observed that (2.1) possibly recovers pretty well the results from simulation, as Tables 2 and 3 show.

Numerical results suggest that the following drawbacks may arise using model (2.1):

- as mentioned before, if N is finite (as in most of the applications), the rules to gather people into groups of size at most L, possibly generate incomplete groups. Thus, the expected number of people assigned to groups of size k, i.e. $N_k = a_k N$ (with $N_1 + \cdots + N_L = N$), is possibly not a multiple of k;
- the model (2.1) relies on the *strong law of large numbers*, which assumes $N \to \infty$. When N is finite the latter consideration, along with the previous item, suggest that for small values of N the model (2.1) may yield inaccurate results (see Table 1);
- Galam's model may even fail to precisely estimate the killing point. We will give an example at the end of Section 4, where the inaccurate killing point prediction by Galam's model causes a wrong computation of the stationary point when $t \to \infty$.

4. A model refinement

Here we propose the model $\mathrm{Gal}_{\mathcal{M}}$, a refinement of formula (2.1) which *partially* complies with the issues listed at the end of the previous section.

Table 4
Average results with $N = 80$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 482, Conv-0 = 18, $L = 6$, $a_1 = a_3 = a_4 = a_5 = 0$, $a_2 = 1/4$, $a_6 = 3/4$. The performance of models (2.1) and (4.3) coincide.

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\|\Delta\|$_Gal	$\|\Delta\|$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8358	0.8358	0.8405	0.0047	0.0047	3.0
3	0.8797	0.8797	0.8812	0.0015	0.0015	3.0
4	0.9238	0.9238	0.9152	0.0086	0.0086	3.1
5	0.9578	0.9578	0.9368	0.0210	0.0210	3.0
6	0.9783	0.9783	0.9497	0.0286	0.0286	3.0
7	0.9891	0.9891	0.9553	0.0339	0.0339	3.1
8	0.9946	0.9946	0.9598	0.0348	0.0348	3.1
9	0.9973	0.9973	0.9618	0.0355	0.0355	3.1
10	0.9986	0.9986	0.9632	0.0354	0.0354	3.1
11	0.9993	0.9993	0.9643	0.0351	0.0351	3.1
12	0.9997	0.9997	0.9649	0.0348	0.0348	3.1
13	0.9998	0.9998	0.9648	0.0351	0.0351	3.1
14	0.9999	0.9999	0.9644	0.0355	0.0355	3.1
15	1.0000	1.0000	0.9640	0.0360	0.0360	3.0
16	1.0000	1.0000	0.9640	0.0360	0.0360	3.1
17	1.0000	1.0000	0.9640	0.0360	0.0360	2.9
18	1.0000	1.0000	0.9640	0.0360	0.0360	3.0
19	1.0000	1.0000	0.9640	0.0360	0.0360	2.8
20	1.0000	1.0000	0.9640	0.0360	0.0360	3.0

Table 5
Average results with $N = 90$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 481, Conv-0 = 19, $L = 6$, $a_1 = a_3 = a_4 = a_5 = 0$, $a_2 = 1/4$, $a_6 = 3/4$. The model (2.1) is almost outperformed by the model (4.3).

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\|\Delta\|$_Gal	$\|\Delta\|$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8358	0.8347	0.8440	**0.0081**	0.0093	2.9
3	0.8797	0.8773	0.8830	**0.0032**	0.0057	3.0
4	0.9238	0.9206	0.9153	0.0085	**0.0053**	3.0
5	0.9578	0.9548	0.9351	0.0227	**0.0197**	2.9
6	0.9783	0.9762	0.9500	0.0284	**0.0262**	3.1
7	0.9891	0.9878	0.9584	0.0307	**0.0294**	3.0
8	0.9946	0.9938	0.9603	0.0343	**0.0335**	3.0
9	0.9973	0.9968	0.9615	0.0358	**0.0353**	3.0
10	0.9986	0.9984	0.9623	0.0364	**0.0361**	3.1
11	0.9993	0.9992	0.9625	0.0368	**0.0366**	3.0
12	0.9997	0.9996	0.9625	0.0372	**0.0371**	3.1
13	0.9998	0.9998	0.9627	0.0371	**0.0371**	3.0
14	0.9999	0.9999	0.9627	0.0372	**0.0372**	3.0
15	1.0000	0.9999	0.9624	0.0375	**0.0375**	3.0
16	1.0000	1.0000	0.9620	0.0380	**0.0380**	3.0
17	1.0000	1.0000	0.9620	0.0380	**0.0380**	2.8
18	1.0000	1.0000	0.9620	0.0380	**0.0380**	2.9
19	1.0000	1.0000	0.9620	0.0380	**0.0380**	3.0
20	1.0000	1.0000	0.9620	0.0380	**0.0380**	3.0

Table 6
Average results with $N = 100$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 500, Conv-0 = 0, $L = 20$, $a_k = 1/20$, $k = 1, \dots, L$. The modified model (4.3) is preferable to (2.1) in the early time steps.

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\|\Delta\|$_Gal	$\|\Delta\|$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.9392	0.9077	0.8868	0.0524	**0.0209**	74.2
3	0.9891	0.9750	0.9513	0.0379	**0.0237**	74.0
4	0.9983	0.9950	0.9819	0.0164	**0.0131**	73.8
5	0.9997	0.9990	0.9945	0.0053	**0.0046**	73.7
6	1.0000	0.9998	0.9984	0.0015	**0.0014**	73.8
7	1.0000	1.0000	0.9995	0.0005	**0.0004**	74.2
8	1.0000	1.0000	0.9999	0.0001	**0.0001**	74.3
9	1.0000	1.0000	1.0000	0.0000	**0.0000**	73.7
10	1.0000	1.0000	1.0000	0.0000	**0.0000**	73.7

We highlight that if N is the size of the overall population, and a_k is the probability that a person is assigned to k-sized groups, then the quantity N_k given by

$$N_k = a_k N,$$

Table 7
Average results with $N = 40$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 423, Conv-0 = 71, $L = 6$, $a_k = 1/6$, $k = 1, \dots, L$. The model (2.1) is almost outperformed by its modified version.

t-step	Gal	$\text{Gal}_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_$\text{Gal}_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8331	0.8206	0.8288	**0.0043**	0.0082	7.2
3	0.8704	0.8445	0.8514	0.0189	**0.0070**	7.5
4	0.9080	0.8707	0.8679	0.0401	**0.0028**	7.2
5	0.9407	0.8976	0.8817	0.0590	**0.0159**	7.4
6	0.9651	0.9229	0.8881	0.0770	**0.0348**	7.5
7	0.9808	0.9447	0.8906	0.0901	**0.0540**	7.3
8	0.9899	0.9619	0.8890	0.1009	**0.0729**	7.5
9	0.9948	0.9746	0.8834	0.1114	**0.0912**	7.4
10	0.9974	0.9834	0.8786	0.1187	**0.1048**	7.5
11	0.9987	0.9894	0.8729	0.1257	**0.1164**	7.4
12	0.9993	0.9932	0.8670	0.1323	**0.1262**	7.5
13	0.9997	0.9957	0.8633	0.1364	**0.1324**	7.3
14	0.9998	0.9973	0.8610	0.1388	**0.1363**	7.7
15	0.9999	0.9983	0.8590	0.1409	**0.1393**	7.6
16	1.0000	0.9989	0.8567	0.1433	**0.1422**	7.4
17	1.0000	0.9993	0.8558	0.1442	**0.1435**	7.4
18	1.0000	0.9996	0.8548	0.1451	**0.1447**	7.3
19	1.0000	0.9997	0.8544	0.1456	**0.1453**	7.4
20	1.0000	0.9998	0.8541	0.1459	**0.1457**	7.5

Table 8
Average results with $N = 60$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 469, Conv-0 = 26, $L = 6$, $a_k = 1/6$, $k = 1, \dots, L$. The model (2.1) is almost outperformed by the model (4.3).

t-step	Gal	$\text{Gal}_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_$\text{Gal}_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8331	0.8200	0.8299	**0.0031**	0.0099	7.3
3	0.8704	0.8436	0.8579	**0.0125**	0.0142	7.6
4	0.9080	0.8700	0.8820	0.0260	**0.0120**	7.5
5	0.9407	0.8974	0.9028	0.0379	**0.0054**	7.7
6	0.9651	0.9237	0.9196	0.0455	**0.0041**	7.4
7	0.9808	0.9464	0.9304	0.0504	**0.0161**	7.5
8	0.9899	0.9644	0.9387	0.0512	**0.0257**	7.6
9	0.9948	0.9773	0.9426	0.0522	**0.0347**	7.5
10	0.9974	0.9860	0.9435	0.0539	**0.0425**	7.5
11	0.9987	0.9915	0.9447	0.0540	**0.0468**	7.3
12	0.9993	0.9950	0.9450	0.0543	**0.0500**	7.4
13	0.9997	0.9970	0.9452	0.0544	**0.0518**	7.6
14	0.9998	0.9983	0.9450	0.0548	**0.0532**	7.8
15	0.9999	0.9990	0.9449	0.0550	**0.0540**	7.5
16	1.0000	0.9994	0.9445	0.0555	**0.0549**	7.6
17	1.0000	0.9997	0.9450	0.0550	**0.0547**	7.6
18	1.0000	0.9998	0.9447	0.0553	**0.0551**	7.3
19	1.0000	0.9999	0.9447	0.0553	**0.0551**	7.4
20	1.0000	0.9999	0.9449	0.0551	**0.0550**	7.5

represents the expected number of people assigned to k-sized groups. As a consequence, the quantity $tail_k$ defined as

$$tail_k = \lfloor N_k \rfloor - \left\lfloor \frac{a_k N}{k} \right\rfloor k, \tag{4.1}$$

represents the number (integral) of people (i.e. a *tail*) assigned to an *incomplete* k-sized group, i.e. $tail_k < k$. Observe that for any k there is at most one incomplete k-sized group, which includes exactly $tail_k$ people. Moreover, at any time step the probability Q_k that an individual is assigned to the incomplete k-sized group containing $tail_k$ elements, is given by

$$Q_k = \frac{tail_k}{a_k N}. \tag{4.2}$$

Therefore, we can consider a new model which encompasses both complete k-sized groups, and incomplete groups with $tail_k$ elements. Based on (4.1)–(4.2), we propose the following refinement of Galam's formula (2.1)

$$P_+(t+1) = \sum_{k=1}^{L} a_k \left[(1 - Q_k) \sum_{j=\left\lfloor \frac{k}{2}+1 \right\rfloor}^{k} C_j^k P_+(t)^j \{1 - P_+(t)\}^{k-j} + Q_k \sum_{i=\left\lfloor \frac{tail_k}{2}+1 \right\rfloor}^{tail_k} C_i^{tail_k} P_+(t)^i \{1 - P_+(t)\}^{tail_k - i} \right]. \tag{4.3}$$

Table 9
Average results with $N = 80$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 474, Conv-0 = 24, $L = 6$, $a_k = 1/6$, $k = 1, \ldots, L$. The model (2.1) is outperformed by its modified version (4.3).

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8331	0.8306	0.8258	0.0073	**0.0048**	7.7
3	0.8704	0.8652	0.8554	0.0150	**0.0098**	7.4
4	0.9080	0.9005	0.8817	0.0263	**0.0188**	7.5
5	0.9407	0.9325	0.9051	0.0356	**0.0273**	7.3
6	0.9651	0.9577	0.9244	0.0407	**0.0333**	7.4
7	0.9808	0.9752	0.9384	0.0424	**0.0369**	7.6
8	0.9899	0.9861	0.9482	0.0417	**0.0379**	7.5
9	0.9948	0.9924	0.9530	0.0418	**0.0394**	7.3
10	0.9974	0.9960	0.9557	0.0417	**0.0403**	7.6
11	0.9987	0.9979	0.9558	0.0429	**0.0421**	7.4
12	0.9993	0.9989	0.9547	0.0446	**0.0442**	7.6
13	0.9997	0.9994	0.9534	0.0463	**0.0460**	7.5
14	0.9998	0.9997	0.9527	0.0472	**0.0470**	7.5
15	0.9999	0.9998	0.9516	0.0483	**0.0482**	7.7
16	1.0000	0.9999	0.9509	0.0491	**0.0490**	7.5
17	1.0000	1.0000	0.9504	0.0496	**0.0496**	7.3
18	1.0000	1.0000	0.9498	0.0502	**0.0502**	7.4
19	1.0000	1.0000	0.9495	0.0505	**0.0505**	7.6
20	1.0000	1.0000	0.9489	0.0510	**0.0510**	7.5

Table 10
Average results with $N = 100$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 486, Conv-0 = 12, $L = 6$, $a_k = 1/6$, $k = 1, \ldots, L$. The model (2.1) is outperformed by the model (4.3).

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8331	0.8274	0.8297	0.0033	**0.0023**	7.3
3	0.8704	0.8591	0.8622	0.0082	**0.0031**	7.6
4	0.9080	0.8928	0.8942	0.0137	**0.0015**	7.4
5	0.9407	0.9248	0.9193	0.0214	**0.0055**	7.6
6	0.9651	0.9515	0.9391	0.0260	**0.0123**	7.6
7	0.9808	0.9709	0.9524	0.0284	**0.0185**	7.6
8	0.9899	0.9835	0.9630	0.0269	**0.0206**	7.2
9	0.9948	0.9910	0.9701	0.0247	**0.0209**	7.5
10	0.9974	0.9952	0.9738	0.0236	**0.0214**	7.4
11	0.9987	0.9975	0.9762	0.0225	**0.0213**	7.4
12	0.9993	0.9987	0.9762	0.0231	**0.0225**	7.6
13	0.9997	0.9993	0.9761	0.0236	**0.0232**	7.4
14	0.9998	0.9996	0.9762	0.0236	**0.0234**	7.5
15	0.9999	0.9998	0.9764	0.0235	**0.0234**	7.5
16	1.0000	0.9999	0.9760	0.0239	**0.0239**	7.7
17	1.0000	0.9999	0.9759	0.0241	**0.0240**	7.3
18	1.0000	1.0000	0.9758	0.0242	**0.0242**	7.5
19	1.0000	1.0000	0.9759	0.0241	**0.0241**	7.4
20	1.0000	1.0000	0.9759	0.0241	**0.0241**	7.7

The first term in square brackets takes into account only the contribution of k-sized groups (as Galam's formula). On the other hand, the second term in square brackets only considers the contribution of incomplete groups with $tail_k$ elements.

Observing (4.3) we note that when $N \to \infty$ our proposal (4.3) coincides with Galam's formula (2.1), since $Q_k \to 0$. In addition, unlike (2.1) the model (4.3) *explicitly* depends on the number of people N, since both Q_k and $tail_k$ depend on N. Thus, if N changes, the probabilities $\{P_+(t)\}$ computed by (4.3) are affected accordingly. Tables 4–13 report the results of the simulations compared with both models (2.1) and (4.3).

From (4.1) and (4.2), if the quantity a_kN/k is integral, then $tail_k = 0$ and $Q_k = 0$. Thus, models (2.1) and (4.3) coincide (an example is given in Table 4). This may be disappointing, since $tail_k = 0$ *does not mean* that no incomplete k-sized groups are formed. Indeed, simulations reveal (see the last column of Table 4) that the average number 'av_tail' of people in the tails (i.e., in incomplete k-sized groups generated by the simulation, with $k = 1, \ldots, L$), may be significant.

Obviously, this drawback appears only in special cases when a_kN/k is integral. In order to avoid the drawback just described, for our model we can simply consider N such that a_kN/k is not integral. For this purpose, in order to carry on a comparison with the results of Table 4, we set $N = 90$ in Table 5. As a result, observe that our proposal seems to be preferable (in Table 5 we could not choose $81 \leq N \leq 89$, since otherwise 80% of N would have not been integral, and a comparison with Table 4 could be possibly meaningless).

Table 11
Average results with $N = 300$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 500, Conv-0 = 0, $L = 6$, $a_k = 1/6$, $k = 1, \ldots, L$. The model (2.1) is often improved by the model (4.3).

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8331	0.8284	0.8324	**0.0006**	0.0040	7.3
3	0.8704	0.8610	0.8678	**0.0025**	0.0069	7.7
4	0.9080	0.8950	0.9028	**0.0052**	0.0077	7.4
5	0.9407	0.9268	0.9336	0.0072	**0.0068**	7.4
6	0.9651	0.9527	0.9574	0.0076	**0.0047**	7.6
7	0.9808	0.9714	0.9740	0.0067	**0.0027**	7.7
8	0.9899	0.9834	0.9848	0.0051	**0.0014**	7.4
9	0.9948	0.9907	0.9914	0.0034	**0.0007**	7.4
10	0.9974	0.9949	0.9951	0.0023	**0.0002**	7.2
11	0.9987	0.9972	0.9975	0.0012	**0.0003**	7.3
12	0.9993	0.9985	0.9987	0.0007	**0.0002**	7.4
13	0.9997	0.9992	0.9992	0.0004	**0.0001**	7.5
14	0.9998	0.9996	0.9996	0.0002	**0.0001**	7.5
15	0.9999	0.9998	0.9998	**0.0001**	0.0001	7.6
16	1.0000	0.9999	0.9999	**0.0000**	0.0001	7.5
17	1.0000	0.9999	1.0000	**0.0000**	0.0001	7.4
18	1.0000	1.0000	1.0000	**0.0000**	0.0000	7.5
19	1.0000	1.0000	1.0000	**0.0000**	0.0000	7.4
20	1.0000	1.0000	1.0000	**0.0000**	0.0000	7.5

Table 12
Average results with $N = 500$, $P_+(0) = 80\%$, 500 runs, Conv-1 = 497, Conv-0 = 0, $L = 6$, $a_k = 1/6$, $k = 1, \ldots, L$. The model (2.1) is outperformed by its modified version (4.3).

t-step	Gal	Gal$_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_Gal$_{\mathcal{M}}$	av_tail
1	0.8000	0.8000	0.8000	0.0000	0.0000	0.0
2	0.8331	0.8330	0.8328	0.0002	**0.0001**	7.5
3	0.8704	0.8700	0.8679	0.0025	**0.0021**	7.7
4	0.9080	0.9072	0.9049	0.0031	**0.0024**	7.5
5	0.9407	0.9398	0.9371	0.0036	**0.0027**	7.6
6	0.9651	0.9642	0.9607	0.0044	**0.0035**	7.7
7	0.9808	0.9800	0.9763	0.0045	**0.0037**	7.6
8	0.9899	0.9894	0.9865	0.0033	**0.0028**	7.3
9	0.9948	0.9945	0.9929	0.0019	**0.0015**	7.6
10	0.9974	0.9972	0.9961	0.0012	**0.0010**	7.5
11	0.9987	0.9985	0.9981	0.0005	**0.0004**	7.4
12	0.9993	0.9993	0.9990	0.0003	**0.0002**	7.5
13	0.9997	0.9996	0.9996	0.0001	**0.0001**	7.5
14	0.9998	0.9998	0.9998	0.0001	**0.0001**	7.6
15	0.9999	0.9999	0.9998	0.0001	**0.0001**	7.2
16	1.0000	1.0000	0.9999	0.0001	**0.0001**	7.7
17	1.0000	1.0000	1.0000	0.0000	**0.0000**	7.6
18	1.0000	1.0000	1.0000	0.0000	**0.0000**	7.5
19	1.0000	1.0000	1.0000	0.0000	**0.0000**	7.6
20	1.0000	1.0000	1.0000	0.0000	**0.0000**	7.5

Table 6 completes the results reported in Table 1 (they are obtained by using the same parameters' settings). Results show that our model often yields smaller errors (bolded results). Again, the column 'av_tail' in Table 6 indicates that there are large tails, which explain our performance.

More generally, on a wide range of numerical tests, the model (4.3) is quite often more accurate than (2.1), in particular when $tail_k$, $1 \le k \le L$, is relatively large. The latter result confirms the theory described. Observe that by simply setting

$$N = 40, 60, 80, 100, 300, 500; \qquad L = 6; \qquad a_k = 1/L, \quad k = 1, \ldots, L,$$

we obtained relatively large values of the tails in our simulations (see Tables 7–13).

So far we have focused on proving that, on average, our proposal is an improvement of Galam's model, rather than detecting possible failures of the latter model. Anyway, we also stressed the performance of our modified Galam's model on a critical numerical test, with $N = 40$, $L = 6$, $a_k = 1/6$, $k = 1, \ldots, 6$. We set $P_+(0) = 28/40 = 0.7$, i.e. $P_+(0)$ is slightly above the fixed point 0.697348 of formula (2.1), which is considered as the killing point by Galam in Ref. [1]. Then, we found out that while Galam's model converges to '1', our model converges to '0', which is closer to the result provided by a simulation taking 1000 runs (see Table 13). In other terms, since $Conv - 1 = 388$ and $Conv - 0 = 587$, it turns out that the threshold $\bar{P}_+$ lies clearly above 0.7.

Table 13
Average results with $N = 40$, $P_+(0) = 70\%$, 1000 runs, Conv-1 = 386, Conv-0 = 589, $L = 6$, $a_k = 1/6$, $k = 1, \ldots, L$. The model (2.1) converges to '1' while our modified model (4.3) converges to '0', according with the results of the simulation.

t-step	Gal	$\mathrm{Gal}_{\mathcal{M}}$	Simul	$\lvert\Delta\rvert$_Gal	$\lvert\Delta\rvert$_Gal$_{\mathcal{M}}$	av_tail
1	0.7000	0.7000	0.7000	0.0000	0.0000	0.0
2	0.7012	0.6896	0.6914	0.0098	**0.0017**	7.4
3	0.7028	0.6750	0.6799	0.0229	**0.0049**	7.4
4	0.7052	0.6540	0.6536	0.0515	**0.0004**	7.3
5	0.7085	0.6236	0.6190	0.0895	**0.0046**	7.3
6	0.7133	0.5789	0.5767	0.1367	**0.0022**	7.4
7	0.7201	0.5125	0.5364	0.1838	**0.0238**	7.5
8	0.7296	0.4157	0.5028	0.2268	**0.0871**	7.5
9	0.7427	0.2851	0.4731	0.2696	**0.1880**	7.4
10	0.7605	0.1440	0.4506	0.3099	**0.3067**	7.4
11	0.7841	0.0469	0.4363	**0.3478**	0.3894	7.7
12	0.8139	0.0107	0.4254	**0.3885**	0.4146	7.4
13	0.8491	0.0021	0.4178	0.4313	**0.4157**	7.2
14	0.8872	0.0004	0.4109	0.4763	**0.4105**	7.0
15	0.9232	0.0001	0.4070	0.5163	**0.4069**	7.5
16	0.9526	0.0000	0.4049	0.5476	**0.4049**	7.3
17	0.9730	0.0000	0.4038	0.5692	**0.4038**	7.5
18	0.9855	0.0000	0.4032	0.5822	**0.4032**	7.4
19	0.9924	0.0000	0.4026	0.5898	**0.4026**	7.6
20	0.9961	0.0000	0.4016	0.5945	**0.4016**	7.4

5. Conclusions and further research

From the results described in the present paper, some issues arise and deserve to be analyzed in future works. Suppose the number N of people is assigned. How should we choose the integer L and the vector $a \in \mathbb{R}^L$ such that the stationary point is reached as soon as possible? We conjecture that both L and a may play a key role.

Moreover, we experienced that in the runs converging to the stationary point '0' (i.e., all agents think '−') the average number of time steps performed is relatively smaller than in the cases of convergence to the stationary point '1' (i.e., all agents think '+'). In this regard we already know that the rule of shifting opinion is slightly biased in favor of the opinion '−' (i.e. towards the stationary point 0), in case of parity. However, we cannot exclude that other specific reasons may yield the latter result.

Now, suppose N is given: how can we modify the information exchange protocol in order to possibly drive the solution towards either the stationary point '1' or '0'? This question turns to be relevant, for example, in marketing, where the successful spreading of a product is a crucial issue.

Finally, let $\mathcal{T}$ be the maximum number of time steps allowed for our model (4.3). Can we predict the probability that a given percentage of agents will be driven to either one opinion or the opposite one within $\mathcal{T}$ time steps? The latter scenario may be of great interest when iterations correspond, for instance, to (expensive) meetings among managers or politicians.

Acknowledgements

The authors wish to thank Paolo Pellizzari for his valuable advice and for the careful reading of the manuscript. The authors also thank the Editor, the referees and the audience of the seminar held at the Department of Applied Mathematics at the University "Ca'Foscari" of Venice.

The second author was supported by INSEAN research program "*VISIR*", and by Program PRIN 20079PLLN7 "*Nonlinear Optimization, Variational Inequalities, and Equilibrium Problems*".

The third author was supported by the Italian PRIN 2007 grant 2007TKLTSR "*Computational markets design and agent-based models of trading behavior*".

References

[1] S. Galam, Modelling rumors: The no plane Pentagon French hoax case, Physica A 320 (2003) 571–580.
[2] E.M. Rogers, Diffusion of Innovations, 5th ed., Free Press, New York, NY, 2003.
[3] D.A. Levy, P.R. Nail, Contagion: A theoretical and empirical review and reconceptualization, Genetic, Social and General Psychology Monographs 119 (1993) 235–284.
[4] F.M. Bass, V. Mahajan, E. Muller, New product diffusion models in marketing: A review and directions for research, Journal of Marketing 54 (1) (1990) 1–26.
[5] T. Lux, The socio-dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distribution, Journal of Economic Behavior and Organization 33 (2) (1998) 143–165.
[6] F. Bass, A new product growth model for consumer durables, Management Science 15 (5) (1969) 215–227.
[7] T. Garber, J. Goldenberg, B. Libai, E. Muller, From density to destiny: Using spatial dimension of sales data for early prediction of new product success, Marketing Science 23 (2004) 419–428.
[8] S. Galam, Contrarian deterministic effects on opinion dynamics: The hung elections scenario, Physica A 333 (2003) 453–460.

[9] S. Galam, F. Jacob, The role of inflexible minorities in the breaking of democratic opinion dynamics, Physica A 381 (2007) 366–376.
[10] M.A. Kamins, V.S. Folkes, L. Perner, Consumer responses to rumors: Good news, bad news, Journal of Consumer Psychology 6 (1997) 165–187.
[11] J. Goldenberg, B. Libai, E. Muller, Talk of the network: A complex systems look at the underlying process of word-of-mouth, Marketing Letters 12 (2001) 209–221.
[12] N. Meade, T. Islam, Modelling and forecasting the diffusion of innovation — A 25-year review, International Journal of Forecasting 22 (2006) 519–545.
[13] S. Galam, B. Chopard, M. Droz, Killer geometries in competing species, Physica A 314 (2002) 256–263.
[14] S. Galam, Global physics: From percolation to terrorism, guerrilla warfare and clandestine activities, Physica A 330 (2003) 139–149.
[15] S. Gekle, L. Peliti, S. Galam, Opinion dynamics in three-choice system, European Physics Journal B 45 (2005) 569–575.
[16] S. Galam, A. Vignes, Fashion, novelty and optimality: An application from Physics, Physica A 351 (2005) 605–619.
[17] S. Galam, Sociophysics: A review of Galam models, 12 Mar. 2008. arXiv:0803.1800v1.
[18] T.L. Roberts, P.B. Lowry, P.D. Sweeney, An evaluation of the impact of social presence through group size and the use of collaborative software on group member voice in face-to-face and computer-mediated task groups, IEEE Transactions on Professional Communication 49 (1) (2006) 28–43.
[19] P.B. Lowry, T.L. Roberts, N.C. Romano Jr., P.D. Cheney, The impact of group size and social presence on small-group communication — Does computer-mediated communication make a difference? Small Group Research 37 (6) (2006) 631–661.
[20] A.R. Dennis, M.L. Williams, Electronic brainstorming: Theory, research and future directions, in: P. Paulus (Ed.), Group Creativity, Oxford University Press, Oxford, 2003.
[21] P. L'Ecuyer, in: Jerry Banks (Ed.), Handbook on Simulation, Wiley, 1997.
[22] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in Fortran 90, vol. 2, Cambridge University Press, 1996.
[23] J.K. Andreazza, L.T. Pinto, Proposition of physiologically plausible neuronal network for the simulation of the primary visual cortex of the Macaque monkey, in: BICS (Brain Inspired Cognitive Systems), Sao Luís, Brasil, 2008.