



HAL
open science

The Leviathan model: Absolute dominance, generalised distrust and other patterns emerging from combining vanity with opinion propagation

Guillaume Deffuant, Timoteo Carletti, Sylvie Huet

► To cite this version:

Guillaume Deffuant, Timoteo Carletti, Sylvie Huet. The Leviathan model: Absolute dominance, generalised distrust and other patterns emerging from combining vanity with opinion propagation. 2012. hal-00678404

HAL Id: hal-00678404

<https://hal.science/hal-00678404>

Preprint submitted on 12 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Leviathan model: Absolute dominance, generalised distrust and other patterns emerging from combining vanity with opinion propagation

Guillaume Deffuant, Timoteo Carletti, Sylvie Huet

March 9, 2012

Abstract

We propose an opinion dynamics model that combines processes of vanity and opinion propagation. The interactions take place between randomly chosen pairs. During an interaction, the agents propagate their opinions about themselves and about other people they know. Moreover, each individual is subject to vanity: if her interlocutor seems to value her highly, then she increases her opinion about this interlocutor. On the contrary she tends to decrease her opinion about those who seem to undervalue her. The combination of these dynamics with the hypothesis that the opinion propagation is more efficient when coming from highly valued individuals, leads to different patterns when varying the parameters. In one of the patterns, absolute dominance of one agent alternates with a state of generalised distrust, where all agents have a very low opinion of all the others (including themselves). We provide some explanations of the mechanisms behind these emergent behaviors and finally propose a discussion about their interest.

1 Introduction

"Again, men have no pleasure (but on the contrary a great deal of grief) in keeping company where there is no power able to overawe them all. For every man looketh that his companion should value him at the same rate he sets upon himself, and upon all signs of contempt or undervaluing naturally endeavours, as far as he dares (...), to extort a greater value from his contemners, by damage; and from others, by the example." T. Hobbes, Leviathan, chapter 13.

In this paper, our goal is to revisit Hobbes thesis using opinion dynamics models and computer simulations. We propose a simple model including a vanity process, inspired from the above citation, in which agents measure themselves in the eyes of the others and retaliate against this judgement. Then, we observe the collective patterns that emerge for different values of the parameters.

Our approach is in the line of many others in the field of social simulation or in sociophysics. It consists in making a few simple assumptions about the rules of interactions between agents and then studying the obtained emerging behaviors. In some of such models, the agents have binary opinions [Galam, 2008; Sznadj-Weron, 2005], while in others the opinions are continuous [Deffuant et al., 2000; Deffuant, 2006; Fortunato, 2004; Huet et al., 2008; Huet

and Deffuant, 2010; Urbig et al., 2008; Lorenz, 2007; F. and Huet, 2010] (see Castellano et al. [2009] for a review). Our model is closer to recent ones which include a set of affinities between agents, leading to emerging networks [Bagnoli et al., 2007; Carletti et al., 2011].

In the proposed model we assume that each agent has a continuous opinion about every other agent and we truncate it to remain between -1 and $+1$. In the initial state, we suppose that the agents don't have any opinion about the others. The agents interact by randomly chosen pairs and two different processes apply. The first one supposes that during any interaction, each agent influences her interlocutor on her opinions about herself, about her interlocutor and about several randomly chosen other agents that she knows. In this propagation, highly valued agents are more influential. The second process represents a vanity effect: an agent likes to be highly valued by the others, thus she increases her opinion on those who value her well. On the contrary she decreases her opinion about those who undervalue on her. These assumptions are inspired by Hobbes, but also by more recent experiments and observations from social-psychologists [Fein, 1997; Leary et al., 2006; Buckley et al., 2004; Srivastava and Beer, 2005; Stephan and Maiano, 2007; Wood and Forest, 2011]. Moreover, we suppose that the access to the opinion of the others is not perfect: people may not express exactly what they think and the listener may misinterpret these expressions. Hence, in the model, the propagated opinions are distorted by some noise.

This paper presents a preliminary analysis, mainly when varying the respective weight of the vanity and opinion propagation. We identified the following main patterns:

- When the vanity is dominating over the opinion propagation, each agent has a positive opinion about herself and she is part of a small cluster of agents connected by strong positive mutual opinions and having negative opinions about all the others. When the influence of vanity gets even stronger, a second category of agents appears, having a negative opinion about themselves and about all the agents of the first category. Moreover, the agents of the second category tend to connect to each other.
- When the opinion propagation is dominating over the vanity, we get consensual reputations: all agents globally share the same opinion about every other agent. When the level of vanity is low, the model exhibits a distribution of the reputations which is widely spread between -1 and $+1$. However, there are more agents of low reputation than of high reputation: this gives the image of a classical hierarchy with a wide basis and progressively shrinking when going up to the top.
- When the weight of vanity gets stronger, a new phenomenon emerges: the absolute dominance of a single (or a couple of) agent of high reputation over the other agents who all have a very low reputation. This dominance is instable, it lasts a variable number of iterations and then it collapses to a generalized distrust where all agents have a negative opinion about all the others (including themselves). After a while, a new episode of dominance takes place.

We firstly describe in details the dynamics of the model. Then we describe the main patterns obtained. We propose theoretical or qualitative explanations of some of these behaviours. Finally, we propose a discussion about these results.

2 The dynamics of the model

This section describes the model in details and particularly its evolution rules representing opinion propagation and vanity.

We consider a set of N agents, each agent i is characterised by her list of opinions about the other agents and about herself: $(a_{i,j})_{1 \leq i, j \leq N}$.

We assume all values of opinions to lie between -1 and $+1$, or *nil* if the agents never met and nobody talked about them yet.

The individuals interact by uniformly randomly drawn pairs (i, j) and at each encounter, we apply both processes: opinion propagation and vanity.

Let us remark that we always keep the opinions between -1 and $+1$, by truncating them to -1 if their value is below -1 after the interaction, or to $+1$ if their value is above $+1$. Moreover, in the following, we consider that one iteration, i.e. one time step $t \rightarrow t + 1$, is N random pair interactions.

2.1 Opinion propagation

Let us assume that agents i and j have been drawn. During an encounter, we suppose that agent j propagates to i her opinions about herself (j), about i , and about k agents randomly chosen among her acquaintances. Moreover, we suppose that if i has a high opinion of j , then j is more influential.

This hypothesis is implemented by introducing a propagation coefficient, hereby denoted $p_{i,j}$, which is based on the difference between the opinion of i about j ($a_{i,j}$) and the opinion i about herself ($a_{i,i}$). It uses the logistic function (σ is a parameter of this function). First of all, if $a_{i,j} = \textit{nil}$, i.e. j is unknown to i , we assume that i has a neutral opinion about j and we set $a_{i,j} \leftarrow 0$. Let us also observe that, because of the initialization, each agent doesn't have any opinion about herself, i.e. $a_{i,i} = \textit{nil}$, before she takes part of a first encounter, thus we also set $a_{i,i} \leftarrow 0$. We thus compute $p_{i,j}$:

$$p_{i,j} = \frac{1}{1 + \exp\left(-\frac{a_{i,j} - a_{i,i}}{\sigma}\right)}. \quad (1)$$

One can easily verify that $p_{i,j}$ tends to 1 when $a_{i,j} - a_{i,i}$ is very positive (i values j higher than herself), and tends to 0 when it is very negative (i values j lower than herself). The influence of j on i is then expressed as follows (ρ is a parameter of the model ruling the importance of opinion propagation):

$$a_{i,i} \leftarrow a_{i,i} + \rho p_{i,j} (a_{j,i} - a_{i,i} + \text{Random}(-\delta, \delta)), \quad (2)$$

and

$$a_{i,j} \leftarrow a_{i,j} + \rho p_{i,j} (a_{j,j} - a_{i,j} + \text{Random}(-\delta, \delta)). \quad (3)$$

Where we denoted by $\text{Random}(-\delta, \delta)$ a uniformly distributed random number between $-\delta$ and $+\delta$, that can be seen as a noise that distorts the perception that i has about j 's opinions. The parameter δ rules the amplitude of this noise.

Moreover, j propagates her opinion about (at most) k of her acquaintances. More precisely, let n_j be the number of acquaintances of j different from i (number of individuals q such that $a_{j,q}$ is not *nil* and $j \neq i$)¹. We choose at random with reinsertion $\min(k, n_j)$ agents among j 's acquaintances (i.e. an acquaintance of j can be selected several times, while other are not selected).

The propagation to i of j 's opinion about q is expressed thus by:

$$a_{i,q} \leftarrow a_{i,q} + \rho p_{i,j} (a_{j,q} - a_{i,q} + \text{Random}(-\delta, \delta)). \quad (4)$$

And this will be repeated for every q selected at random.

In the interaction, we apply the influence of j on i and then the reciprocal one of i on j .

¹We also tested without checking that the acquaintance is different from i , and did not notice any change in the emergent patterns of the model.

2.2 Vanity dynamics

The second ingredient of the model is the dynamics representing agent’s vanity. This dynamics expresses that agents tend to reward the agents that value them more positively than they value themselves and to punish the ones that value them more negatively than they value themselves.

We assume that the modification of i ’s opinion about j is simply linear with the difference between the opinion of i about herself and the opinion of j about i (slightly modified randomly):

$$a_{i,j} \leftarrow a_{i,j} + \omega(a_{j,i} - a_{i,i} + \text{Random}(-\delta, \delta)). \quad (5)$$

If i has a lower (resp. higher) self opinion than her perception of the opinion j has about her (i), then $a_{i,j}$ is increased (resp. decreased). Parameter ω rules the importance of the vanity dynamics. We also assume that the intensity of the vanity effect is independent from the opinions. Indeed, it is possible that one forgives more easily a disappointment coming from a loved person because precisely of this affection. But on the other hand, it is well known that hatred can also easily come from disappointed love. Since there is no clear intuition in one way or the other, we made the simplest assumption that the vanity effect is linear with the difference of opinions.

2.3 Summary

The Algorithm 1 describes one iteration, i.e. one time step: N random pairs of individuals are drawn, with reinsertion, and we suppose that each individual influences the other during the encounter. Algorithm 2 codes the interaction with the two aspects of the dynamics: opinion propagation and vanity.

Algorithm 1: Iteration

```
for  $N$  times do
   $i \leftarrow \text{Random}(1, N)$  ;
   $j \leftarrow \text{Random}(1, N, \neq i)$  ;
  Interaction( $i, j$ ) ;
  Interaction( $j, i$ ) ;
end
```

Summarizing the parameters of the model are thus:

- N , number of agents (in the experiments presented in the next sections, we fixed $N = 40$);
- ρ , ruling the intensity of the opinion influence;
- ω , ruling the intensity of the vanity;
- k , number of acquaintances about which the pair of agents discuss in the opinion influence (in the experiments presented in the next sections, we fixed $k = 10$);
- δ , intensity of noise perturbing the evaluation of other’s opinions (in the experiments presented in the next sections, we fixed $\delta = 0.2$);
- σ , ruling the slope of the logistic function determining the propagation coefficients (in the experiments presented in the next sections, we fixed $\sigma = 0.3$).

Algorithm 2: Interaction(j, i)

```
if  $a_{i,i} = nil$  then  $a_{i,i} \leftarrow 0$ 
if  $a_{i,j} = nil$  then  $a_{i,j} \leftarrow 0$ 
if  $a_{j,i} = nil$  then  $a_{j,i} \leftarrow 0$ 
if  $a_{j,j} = nil$  then  $a_{j,j} \leftarrow 0$ 
 $p_{i,j} \leftarrow \frac{1}{1 + \exp\left(-\frac{a_{i,j} - a_{i,i}}{\sigma}\right)}$  ; // Computing propagation coefficient
 $a_{i,i} \leftarrow a_{i,i} + \rho p_{i,j} (a_{j,i} - a_{i,i} + \text{Random}(-\delta, \delta))$  ; //  $j$  propagates  $a_{j,i}$ 
 $a_{i,j} \leftarrow a_{i,j} + \rho p_{i,j} (a_{j,j} - a_{i,j} + \text{Random}(-\delta, \delta))$  ; //  $j$  propagates  $a_{j,j}$ 
for min( $k, n_j$ ) times do ; //  $n_j$  number of individuals known by  $j$ 
     $q \leftarrow \text{Random}(\text{KnownAgents}(j), \neq i, \neq j)$  ; //  $q$  random known by  $j$ 
    if  $a_{i,q} = nil$  then  $a_{i,q} \leftarrow 0$ 
     $a_{i,q} \leftarrow a_{i,q} + \rho p_{i,j} (a_{j,q} - a_{i,q} + \text{Random}(-\delta, \delta))$  ; //  $j$  propagates  $a_{j,p}$ 
end
 $a_{i,j} \leftarrow a_{i,j} + \omega (a_{j,i} - a_{i,i} + \text{Random}(-\delta, \delta))$  ; // Vanity dynamics,  $i$  modifies  $a_{i,j}$ 
```

2.4 Representations of the population state

We use two different representations of the population state:

- **Matrix representation:** the opinion list of each agent is represented as the row of a $N \times N$ square matrix. The element $a_{i,j}$ from line i and column j is the opinion of agent i about agent j . We use colours to code for the opinions: blue for negative and red for positive opinions with light colours meaning that the absolute value is close to 0. This representation provides all the information about the state of the population at a given time step, but it is sometimes difficult to interpret.
- **Network representation:** we represent the agents as nodes of a network, in a 2D space and we draw links between two agents, only when one of the agents has a positive opinion about the other. The position in the 2D space of the nodes representing the agents is obtained by a dynamical algorithm, which at each interaction moves the nodes in order to get a distribution where the distances between the nodes are as close as possible to the values of a simple function depending on the sum of the relative agents' opinions. The function yields a large value when two agents have negative opinions about each other, and reciprocally, a small value when their opinions about each other are positive. When the sum of the opinions is close to 0, the function yields a medium value. The colour of the links is yellow when the link is close to 0 and it get close to red when the value of the link is close to 1. This representation does not show all the information about the state, because we do not represent the negative links nor the assymetry of the links. However, some features of the emergent structures appear more easily with this representation.

In most cases, we propose both representations for the same state.

3 Vanity dominating patterns

We identified two main patterns when parameter ρ (weighting the opinion propagation) is small compared to 1 and parameter ω (weighting the vanity) is significantly higher. The

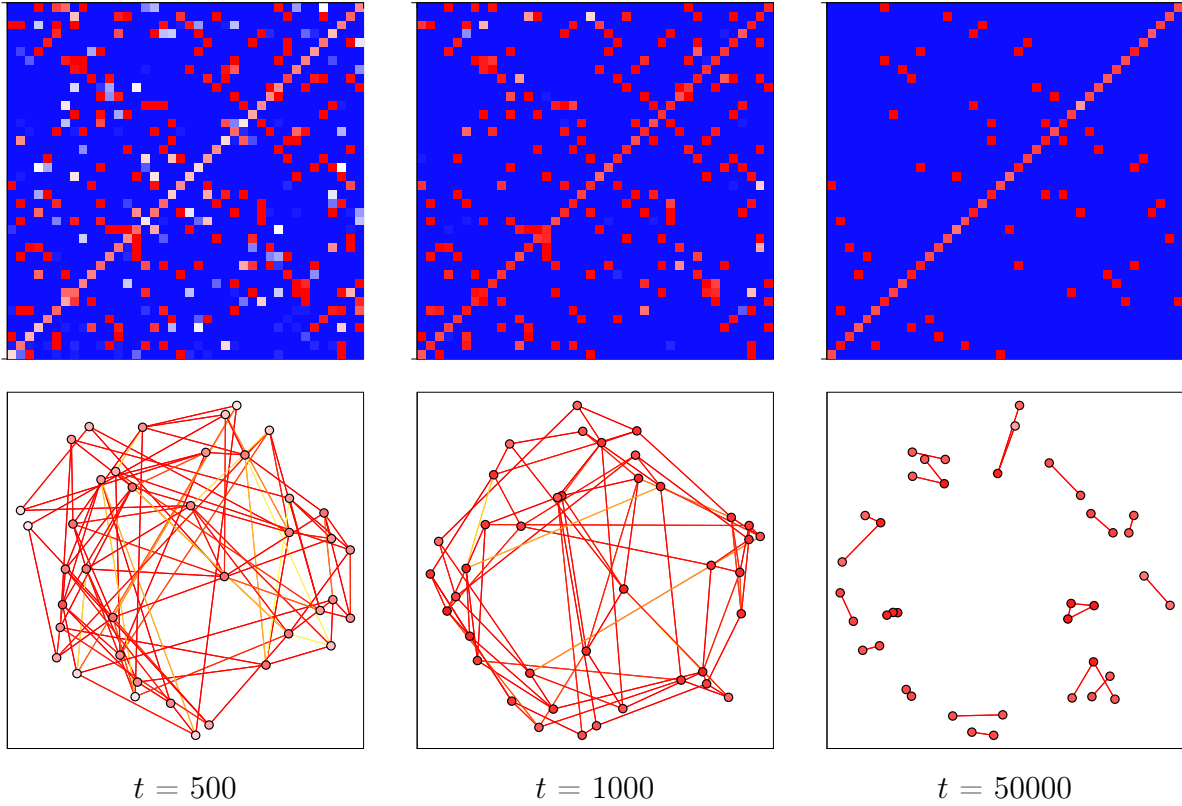


Figure 1: Example of pattern1. Parameters are: $\omega = 0.3$ and $\rho = 0.02$. See Section 2.4 for general explanation about the representations.

following two sections describe these patterns.

3.1 Pattern 1. Small sets of strongly positively connected individuals, with positive self value

Figure 1 shows, using, both matrix and network representations, three states of the model for $\omega = 0.3$, $\rho = 0.02$ (the other values of the parameters are given in section 2.3. The model converges rapidly to a stable configuration, in which the opinion matrix is symmetric, with only extreme values (-1 or $+1$). Each agent separates the population between two groups: the ones she hates and the ones she loves. The first group is much larger than the second (the average population opinion is less than -0.8). Another feature is striking: the opinions on the diagonal are significantly positive: the agents have a good opinion of themselves (though not as good as their opinion about their loved ones). Looking at the network representation of the stationary state (iteration 50000), we observe the presence of small connected sets made by 2 or 3 individuals. The sets tend to be located far from each other (on a circle), indicating that each connected set has a negative opinion about the others. At iterations 500 and 1000, the network of positive connections is much more dense. Many positive connections are progressively destroyed by the interactions before reaching the stationary state.

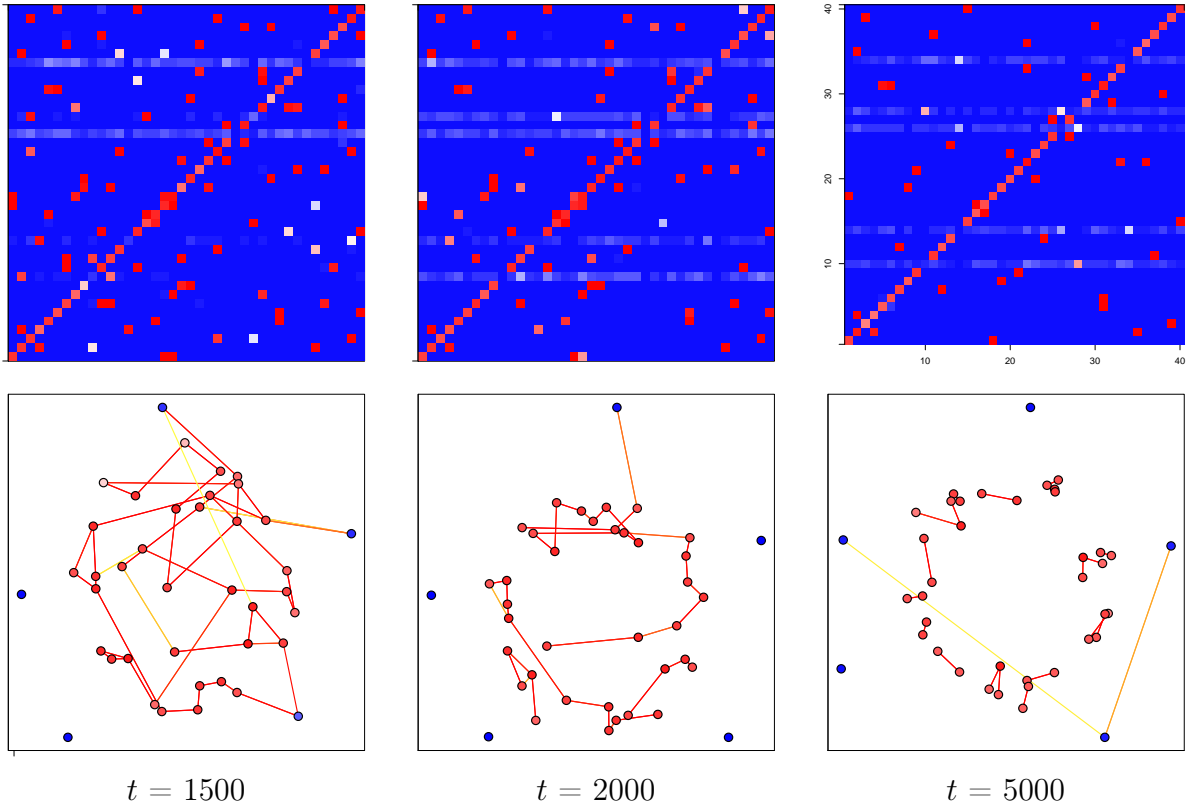


Figure 2: Example of pattern2. $\omega = 0.3$ and $\rho = 0.04$. Several individuals, with a low opinion of themselves appear. See section 2.4 for general explanation about the representations.

3.2 Pattern 2. Isolated agents, negatively valuing themselves, superposed to pattern1

When the weight of the opinion propagation increases, the previous pattern is modified as shown on Figure 2, where $\rho = 0.04$, while all the other parameters values have been fixed as previously. We observe that the matrix is no longer symmetric, a few horizontal lines appear with a majority of slightly negative values, the value of the diagonal for these lines being negative. Moreover, these lines are not stable, their number and positions vary in time. When looking at the network representation, we observe that the previous ring of positively connected sets is surrounded by a second ring of 5 blue nodes (hence with a negative opinion of themselves). Such a blue node appears when a connected agent gets a negative self opinion, and then breaks her links with the others (see the difference between $t = 1500$ and $t = 2000$). Moreover, the second ring of agents tends to establish connections between them (for instance for $t = 5000$). Actually, when several of them connect, they tend to increase their self opinion, and they may come back to the first ring (not shown on the figure). Meanwhile, agents from the first ring can get a negative opinion of themselves, break all their connections and move to the second ring. Globally the number of agents of each ring fluctuates around a constant average value over the time.

3.3 First explanations of patterns 1 and 2

In order to better understand these patterns, we simulate the model with the vanity dynamics alone, i.e. $\rho = 0$, starting with all agents not knowing each other, hence with all the opinions

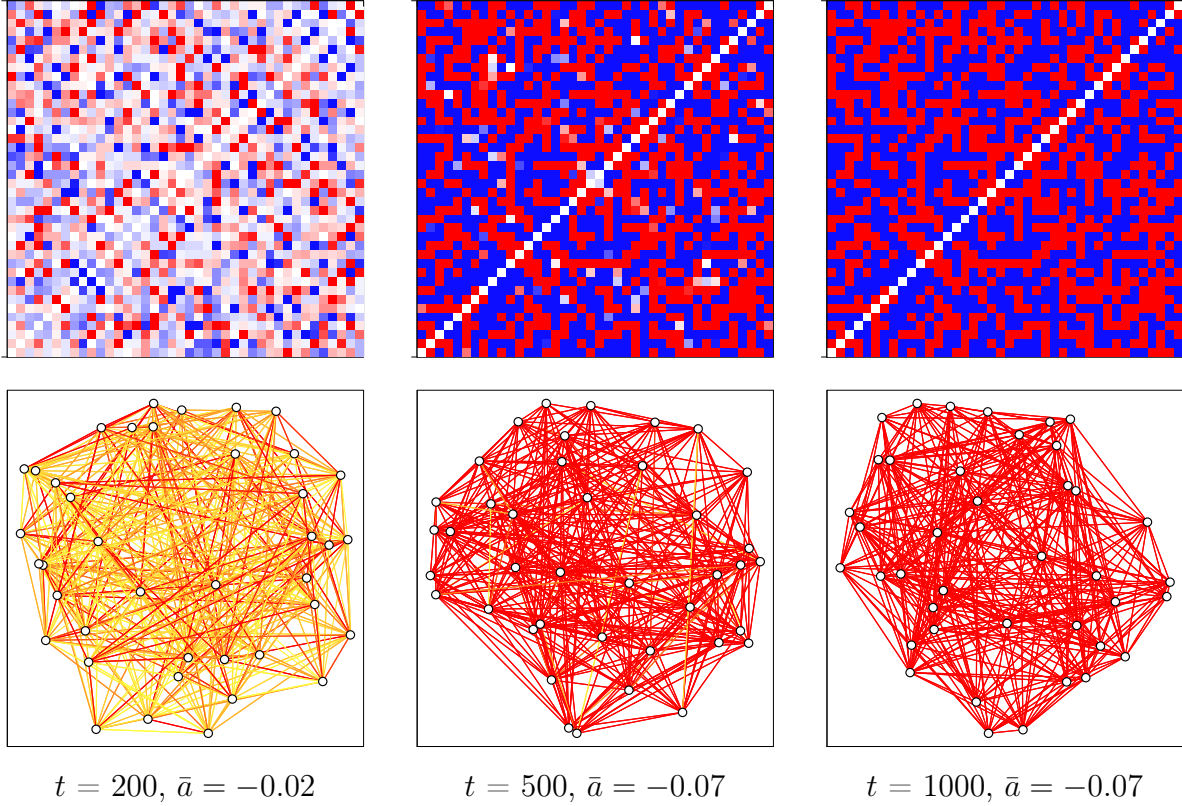


Figure 3: Vanity only. $\rho = 0$, $\omega = 0.4$. The matrix remain symmetric and each agent splits the population between friends and foes. See section 2.4 for general explanation about the representations.

set to *nil*. We get the evolution represented on Figure 3. After relatively few time steps, all the opinions become either -1 or $+1$, except the self opinions which remain *nil* (the diagonal of the matrix representation remains white). Each agent separates the population into two groups of almost equal size: the ones she hates and the ones she loves. We also note that the diagonal is a symmetry axis of the matrix: if I love you, you love me, or if I hate you, you hate me. Note finally that the average of all opinions over the whole population is close to 0.

We now propose some analytical explanations of the above result. By definition of the vanity process, and because $a_{i,i} = a_{j,j} = 0$, when agents i and j meet, the opinions evolve as follows:

$$\begin{cases} a_{i,j}(t+1) \leftarrow a_{i,j}(t) + \omega(a_{j,i}(t) + \text{Random}(-\delta, \delta)) \\ a_{j,i}(t+1) \leftarrow a_{j,i}(t) + \omega(a_{i,j}(t) + \text{Random}(-\delta, \delta)). \end{cases} \quad (6)$$

From these equations, we can derive the following points:

- If we have, for a couple (i, j) with $i \neq j$, at a given iteration t , $a_{i,j}(t) > \delta$ and $a_{j,i}(t) > \delta$, then $a_{j,i}(t) + \text{Random}(-\delta, \delta) > 0$ and $a_{i,j}(t) + \text{Random}(-\delta, \delta) > 0$. Hence, in this case, after iteration t , both $a_{i,j}$ and $a_{j,i}$ keep increasing until they are truncated to 1.
- Similarly, if $a_{i,j}(t) < -\delta$ and $a_{j,i}(t) < -\delta$, then both $a_{i,j}$ and $a_{j,i}$ keep decreasing afterwards until they are truncated to -1 .
- If $a_{i,j}(t) < -\delta$ and $a_{j,i}(t) > \delta$, then at the next steps $a_{i,j}(t)$ will increase and $a_{j,i}(t)$ will decrease. Of course we get the same when interverting i and j .

This allows us to describe the evolution of the reciprocal opinions of each couple (i, j) in two

phases:

- In a first phase, when the absolute values of $a_{i,j}$ and $a_{j,i}$ are smaller than δ , the effect of randomness is dominating, and the opinions tend to be close to random walks. When both opinions get close to δ , these random walks are biased towards common growth, when they both get close to $-\delta$, they are biased towards common decrease, and if their signs are different, the opinions tend both to go to 0.
- As a result of these random processes each opinion couple $(a_{i,j}, a_{j,i})$ necessarily ends up either by being both higher than δ or both lower than $-\delta$, with an equal probability. This leads to the opinions couples being both at $+1$ or -1 in the end.

This analytical study helps to formulate qualitative explanations of pattern 1. First, we make the hypothesis that the dynamics of opinion propagation is negligible except on the diagonal of the matrix, because out of the diagonal the vanity dynamics are dominating. The evolution of the opinions outside the diagonal is given by the vanity equations:

$$\begin{cases} a_{i,j}(t+1) \leftarrow a_{i,j}(t) + \omega(a_{j,i}(t) - a_{i,i}(t) + \text{Random}(-\delta, \delta)) \\ a_{j,i}(t+1) \leftarrow a_{j,i}(t) + \omega(a_{i,j}(t) - a_{j,j}(t) + \text{Random}(-\delta, \delta)). \end{cases} \quad (7)$$

If all the values of the diagonal are positive, we see from these equations that this tends to decrease the values of $a_{i,j}$ compared with the case of $\rho = 0$, which explains why these values are more frequently negative. It remains to be explained why the values of the diagonal are positive. Suppose agent j propagates her opinion about i to i herself, we have:

$$a_{i,i} \leftarrow a_{i,i} + p_{i,j}\rho(a_{j,i} - a_{i,i} + \text{Random}(-\delta, \delta)). \quad (8)$$

We note that $p_{i,j}$ is high when $a_{j,i}$ is high, i.e. when $a_{j,i} - a_{i,i}$ tends to be positive. On the contrary, $p_{i,j}$ is close to 0 when $a_{j,i}$ is low, i.e. when $a_{j,i} - a_{i,i}$ tends to be negative. Therefore, the propagation of opinions favours the positive contributions to the diagonal, explaining why this diagonal is positive.

Qualitatively, the agents love those who flatter their ego and believe more those they love. The result is that they tend to have a positive self opinion. These positive self opinions tend to shift the other opinions to the negative side, because the agents have a self opinion which is higher than the average opinion about them. Indeed, the agents are disappointed most of the time by the opinion of others and they decrease their opinion about them by vanity. In return, these agents do the same.

After a while, we get an equilibrium where roughly the number of -1 opinions weighted by the propagation coefficient balances the number of $+1$ opinions weighted by the propagation coefficient. However, if this self opinion of equilibrium is such that $a_{i,i} > 1 - \delta$, then the random fluctuations can decrease $a_{i,j}$, even if $a_{i,j} = 1$. In this case, because of these fluctuations, some couples $(a_{i,j}, a_{j,i})$ that converged to $+1$, enter in negative retaliation loops and finally converge to -1 . This leads $a_{i,i}$ to a smaller equilibrium value. The process continues until $a_{i,i} < 1 - \delta$, for all i .

We can also get a qualitative understanding of pattern 2. When ρ increases, the fluctuations on the diagonal values increase as well. Even with a single value of j such that $a_{i,j} = 1$, we can have, at some time steps, $a_{i,i} > 1 - \delta$. In this case, even the last support of i can be lost. To summarise, when the value of the ego becomes too high, the agent has higher chances to loose all her friends. Once the agent becomes isolated, she progressively loses her high self opinion because all the other agents are sending her messages of negative value. This explains why these agents form a second ring in the network representation.

4 Patterns with strong opinion propagation

4.1 Pattern 3. Hierarchy of commonly agreed reputations

When the opinion propagation gets stronger than in the previous examples, all agents tend to have the same opinion about each other agent, hereby called for convenience the reputation r_i of agent i . This is shown for instance on Figure 4 obtained using $\omega = 0.2$ and $\rho = 0.5$. In the matrix representations, this is visualized by cells with similar colors in the same column. As explained more formally later, this agreement of opinions takes place because of the strength of the averaging dynamics in the opinion propagation. The random noise added to the propagation introduces some random fluctuations of the opinions around the average value of the reputation and also some fluctuations of this average value.

Moreover, if we observe the distribution of the reputations, it appears that the number of agents is decreasing with the reputation. This gives an image of a hierarchy of reputations with a large basis and a progressively shrinking number of agents while going up to an exclusive elite. It can be observed on the matrix representation that the members of this elite (the red columns) are continuously changing: the members of high reputation at a given moment can drop later on to the basis of the hierarchy and conversely an agent of the lowest reputation can get to the highest elite after a while.

Moreover, the distribution itself is not fully stable. We observe that, from times to times a very strong dominance of a few agents that have a reputation close to 1, whereas all the other reputations are close to -1. The dominant agents generally rapidly loose their high reputation which decrease to medium levels (0.5 or even a bit lower), and then the pattern of a stronger dominance appears again.

4.2 Pattern 4. Short periods of absolute dominance and long periods of general distrust

We consider now the pattern with the vanity playing a more important role ($\omega = 0.5$ and $\rho = 0.3$). In this case we get a different pattern, with periods of strong dominance by one or two agents, followed by periods of generalized distrust, all the agents having very negative opinions of all the others and also about themselves. Figure 5 proposes the matrix and network representations of six states around the first peak of dominance. From iteration 6500 to iteration 7500, two individuals, shown as red columns in the matrix representation whereas all the other columns are blue and in red nodes located in the centre of the network representations, are valued highly positively (red links) by all the others which are colored in blue and located far from the circle center. After the dominance period, there are no more positive links, and all the nodes are blue (negative self value).

Other dominance episodes start with only one dominant individual. In this case, the dominance tends to last longer and it can lead to progressive hierarchy that are similar to the ones observed in pattern 3. The dominance ends when a second dominant individual emerges and reaches the same level of reputation as the first dominant one. This situation is very instable and it leads rapidly to the generalised distrust. This case is studied in more details later.

To characterize the distribution of reputations, we define a dominance indicator D based on the quantities \bar{a} , r_M and r_m respectively the average, the maximum and the minimum of the reputations:

$$D = \left(r_M - \bar{a} - \frac{r_M - r_m}{2} \right). \quad (9)$$

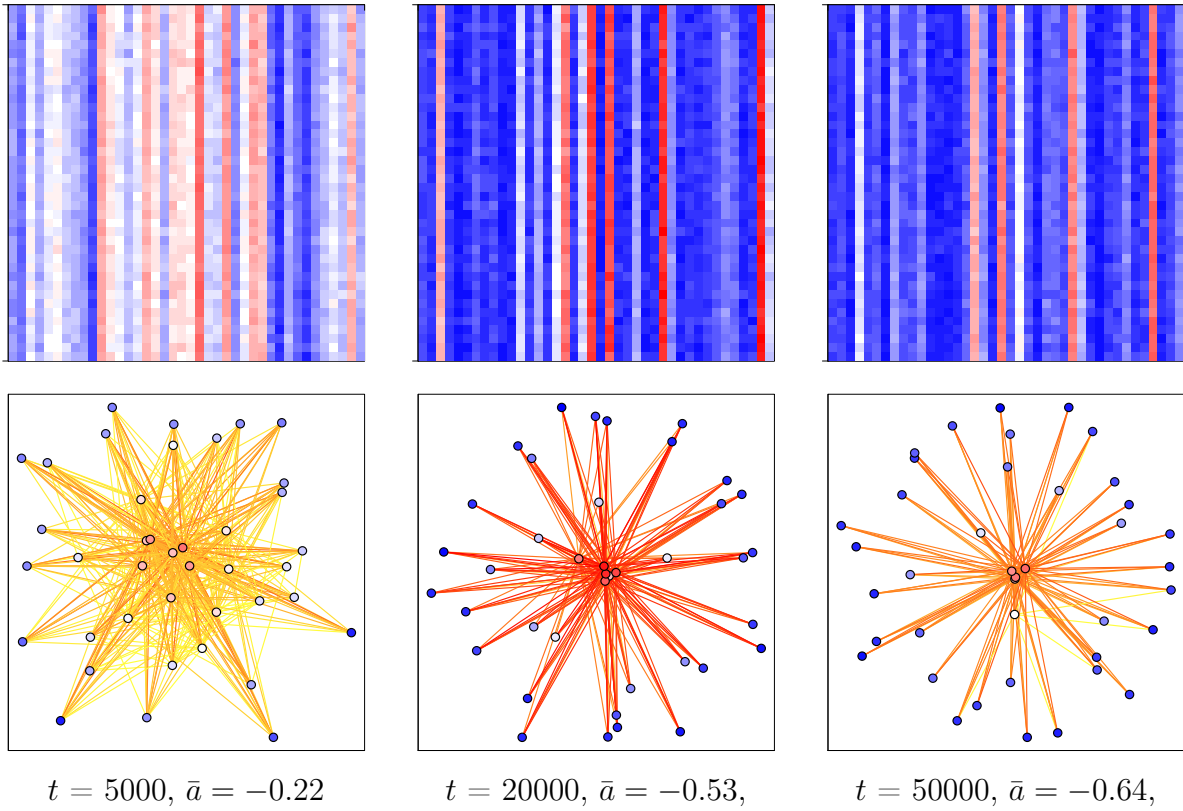


Figure 4: Example of pattern3: stable hierarchy. $\omega = 0.2$, $\rho = 0.5$, $N = 40$, $\delta = 0.4$, $k = 10$, $\sigma = 0.3$. See section 2.4 for general explanation about the representations.

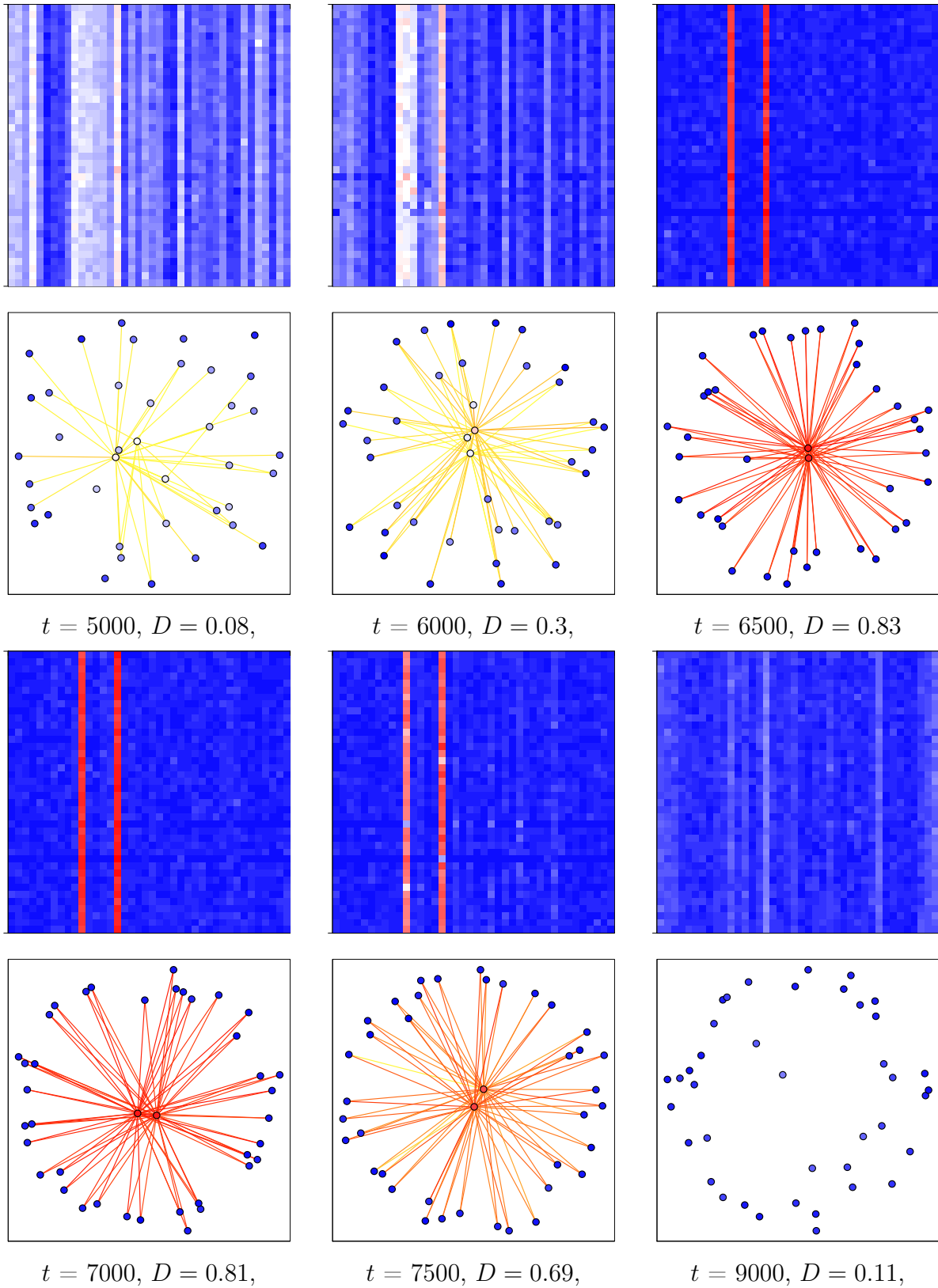


Figure 5: Example of pattern4. Study around the first peak of the dominance indicator for $\omega = 0.5$, $\rho = 0.3$, between iterations 5000 and 9000. In a first phase, two agents reinforce each other to dominate the whole population and in the second phase they compete and destroy each other's reputation. See section 2.4 for general explanation about the representations.

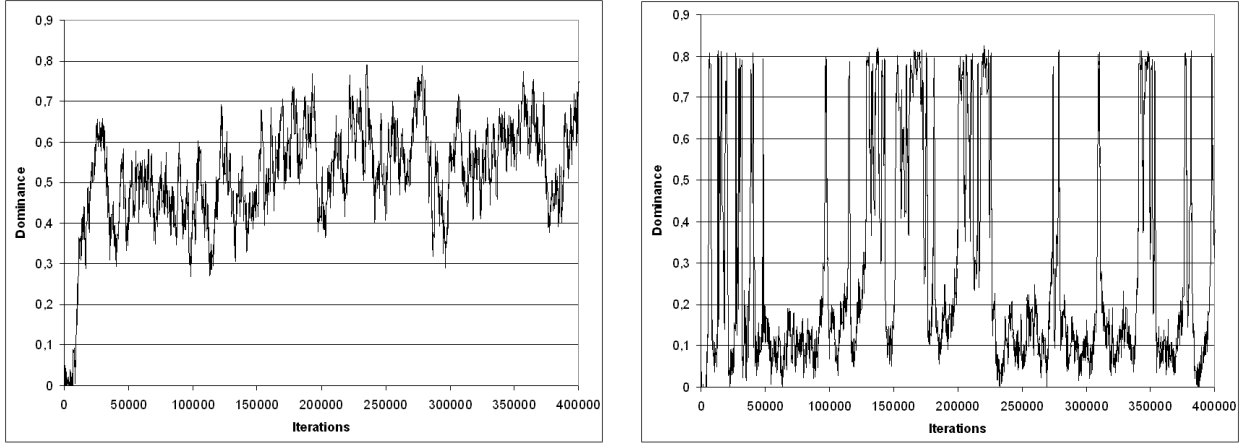


Figure 6: Evolution of the dominance indicator ($\times 100$) characterizing the distribution of reputations (see definitions of the indicators in the text). (Left panel) $\omega = 0.2$ and $\rho = 0.5$ (opinion influence clearly dominating), the fluctuations of the indicators are much smaller than for (right panel) $\omega = 0.5$ and $\rho = 0.3$ (vanity playing a bigger role).

The rationale behind this indicator can be understood when considering extreme cases for the reputation distribution:

- When the reputations are uniformly distributed around their average \bar{a} (for instance when the distribution is uniform), the dominance indicator is close to 0, because in this case the distance from the highest reputation r_M to the average \bar{a} is around half of the total extension of the reputation distribution (this extension equals 2).
- When the distribution of reputations is very narrow (all reputations are almost the same), then the dominance is close to 0 as well.
- When one individual (or a small number of them) has a very high reputation (say +1) and all the others have a very negative reputation (say -1), then the dominance indicator is large and close to 1.

To summarise, the indicator is close to 1 when one individual has the maximum reputation and the others have the lowest one, the indicator is close to 0 when there is no clear dominance.

Figure 6 shows the evolution of the dominance for two sets of parameters : $\omega = 0.2$ and $\rho = 0.5$ (left panel) and for $\omega = 0.5$ and $\rho = 0.3$ (right panel). The graphs suggest some comments:

- For $\omega = 0.2$ and $\rho = 0.5$, we note that the dominance is rapidly growing and then shows fluctuations between 0.3 and 0.8. The peaks that are higher than 0.7 are episodes of strong dominance by two agents.
- For $\omega = 0.5$ and $\rho = 0.3$, the graph has a different shape: we note very strong fluctuations, with several major peaks up to more than 0.8, indicating short periods of very high dominance by one or two agents. When this dominance lasts longer (for instance after iteration 150 000 and after iteration 200 000), it indicates that a single individual is dominating. We also observe that the dominance is very low between these peaks (less than 0.2) indicating that all the reputations are almost the same, and in this case very negative.

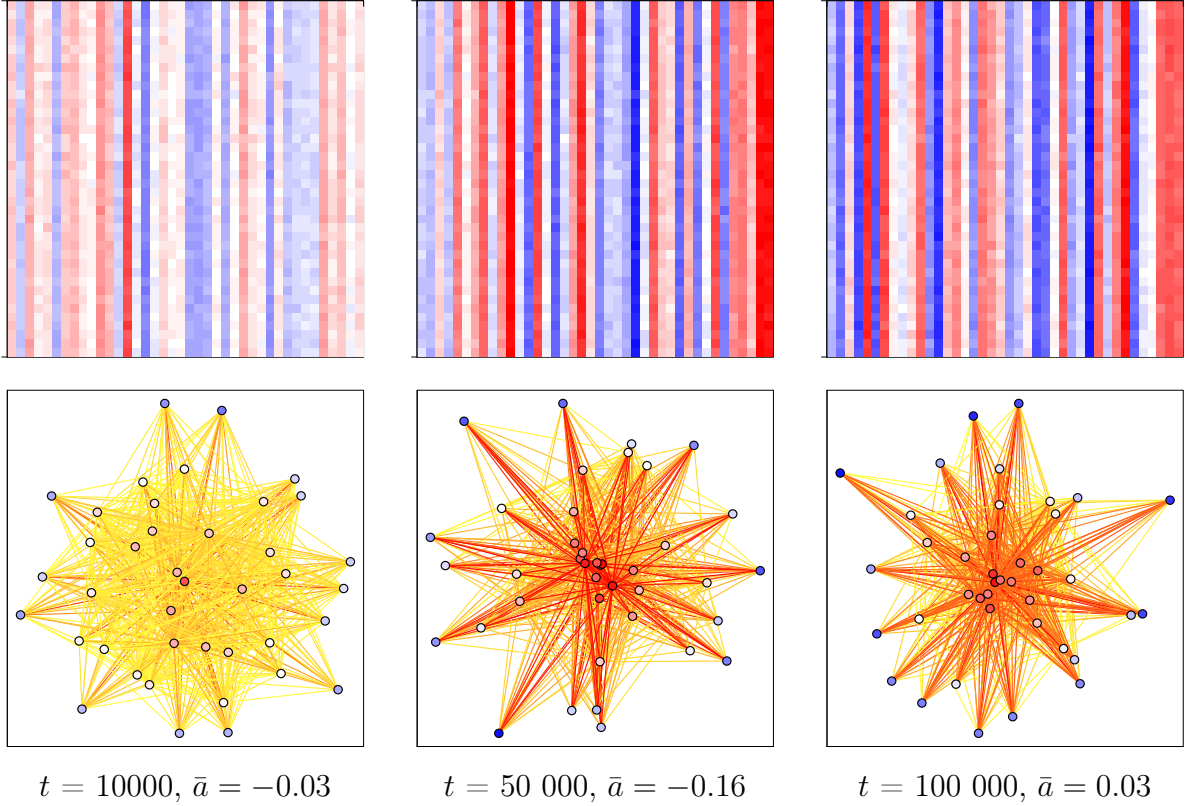


Figure 7: Evolution of the opinions with the dynamics of opinion propagation only ($N = 40$, $\omega = 0$, $\rho = 0.3$).

4.3 First explanations of patterns 3 and 4

4.3.1 Study of the case $\omega = 0$

Studying the model with the opinion influence only ($\omega = 0$) can give us some clues. Figure 7 shows the corresponding pattern, that suggests the following observations:

- Like in patterns 3 and 4, all the agents have similar opinions about each agent, leading to matrix representations with columns of almost homogeneous colours (the reputation r_i of agent i).
- Unlike in patterns 3 and 4, after a while, the values r_i do not show a strong dominance. The values r_i are not stable over time, they can take any value in the segment $[-1, +1]$ if the simulation is long enough.

In the case $\omega = 0$, it is easier to show analytically why the columns of the matrix tend to be homogeneous.

Let us assume i and j meet, the propagation of opinion process implies:

$$\begin{cases} a_{i,j}(t+1) &= a_{i,j}(t) + \rho p_{i,j} (a_{j,j}(t) - a_{i,j}(t) + \delta_{j,j}) \\ a_{j,j}(t+1) &= a_{j,j}(t) + \rho p_{j,i} (a_{i,j}(t) - a_{j,j}(t) + \delta_{i,j}) \end{cases} \quad (10)$$

Let us introduce the variables $D_{i,j}(t) = a_{i,j}(t) - a_{j,j}(t)$, hence

$$D_{i,j}(t+1) = D_{i,j}(t) (1 - \rho(p_{i,j} + p_{j,i})) + \Delta, \quad (11)$$

where $\Delta = \rho p_{ij} \delta_{j,j}(t) - \rho p_{ji} \delta_{i,j}(t)$ is the stochastic contribution. By definition $p_{ij} + p_{ji} \in [0, 2]$ and it can be equal to zero only $a_{ii} - a_{ij} \gg 1$ and $a_{jj} - a_{ji} \gg 1$. So neglecting for a while the stochastic term and defining $\theta = (1 - \rho(p_{ij} + p_{ji}))$ we have:

$$|D_{i,j}(t+1)| \leq \theta |D_{i,j}(t)|, \quad (12)$$

because, if we suppose $\rho < 0.5$, we have $\theta < 1$ and we get that as t increases $D_{i,j}(t+1)$ goes to zero. This is the required result.

Now, with the addition of the stochastic part, the dynamics of the reputations seems to be a random walk between -1 and $+1$. However, when computing the distribution of the reputations over 50000 iterations in a sliding window, we observe that this distribution is not perfectly uniform. Figure 8 on the left shows the distribution of the self opinions (which are close to the reputations) and we observe that the average value tends to be slightly negative, and the distribution has two maxima close to its extremes (around -0.7 and around $+0.8$), with a minimum between these maxima around 0.3 .

Moreover, we observe on Figure 8 on the right that the average self-opinion tends to be slightly higher than the reputation, except when the self-opinion of the agents are close to the extremes. The values of the extremes are due to side effects: the reputation tends to fluctuate more slowly than the self opinion. The average positive bias for the self opinion is more interesting. It is due to the propagation coefficient which tends to be higher when the self-opinion gets higher. Indeed, because of this difference, when an agent self-opinion is higher than her reputation, the others have less influence on the self-opinion than when the self-opinion is lower than the reputation (everything else being equal). However, the effect of this average difference between the self opinion and the reputation depends on the value of the agent's reputation:

- If the agent's reputation (and hence her self-opinion) is among the lowest, then she is very sensitive to the influence of most of the others that she values more than herself (they have thus a high propagation coefficient). Therefore, when the agent self-opinion is higher than her reputation (which statistically takes place more often than the opposite), the agent's self-opinion tends to follow her reputation (the average opinion of the others about herself), thus it decreases. This explains a tendency towards negative values.
- If the agent's reputation (and hence her self-opinion) is among the highest, then on average she tends to impose her self opinion to the others, because her propagation coefficient is high in the exchanges. Thus, on the contrary, when her self-opinion is higher than her reputation (statistically the most frequent), she tends to increase her reputation towards her self opinion.

To summarise, the highly valued individuals tend to lead the other's opinions and, with the statistical bias for a self-opinion higher than the reputation, they tend to increase their reputation. This is the contrary for the badly valued individuals who tend to naturally decrease their self-opinion, only by the effect of the propagation coefficient. Nevertheless, these are statistical tendencies, which take place stochastically and there are random movements in the opposite directions. This explains the shape of the reputation distribution. Below a threshold of the self opinion, the reputations and self-opinions tend to be biased towards negative values, and above towards positive values. The value of this threshold depends on parameter σ determining the propagation coefficient and also on k the number of agents about whom the opinions are propagated during the encounters. Indeed, this number has an impact on how the agents propagate their opinion about the others.

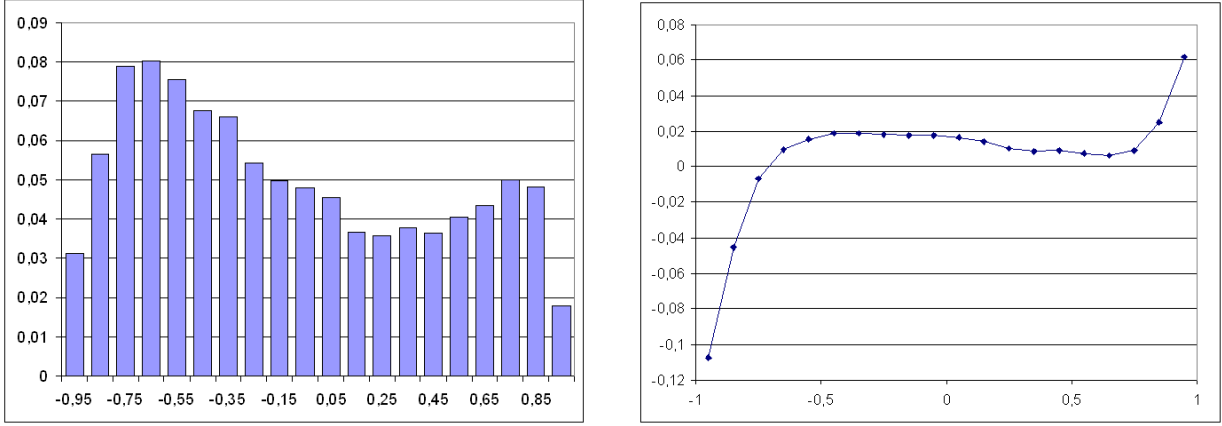


Figure 8: Graphs for $\omega = 0$, $\rho = 0.3$. On the left: the density of self opinion averaged on 50000 iterations. On the right: the average difference between the self opinion and the reputation $(a_{i,i} - r_i)$.

4.3.2 Analysis of patterns 3 and 4

These observations are useful to understand the processes behind the emergence of patterns 3 and 4. First note that the vanity process enhances the tendency of self-opinions to be higher than the reputations. Indeed, the small statistical positive bias for self opinion that is due to the opinion propagation leads, on average, the agents to consider themselves as (slightly) undervalued by the others, thus they devalue them by vanity. This is very similar to the process that we observed in pattern 1, but it is slower because of the averaging effect of the opinion propagation. This explains the tendency to the negative opinions in both patterns 3 and 4, with its extreme expression with the generalised distrust of pattern 4.

But the generalised distrust of pattern 4 also generates episodes of dominance that we now try to explain qualitatively:

- There are fluctuations of the reputations during the generalised distrust. When an agent's reputation (and thus self-opinion) becomes higher enough than the others, it becomes less sensitive to the opinion propagation from the others. In this case, the opinions of this leading agent become more strongly driven by the vanity than by the opinion propagation. In particular, this agent can establish loops of positive reinforcements with a few other agents that are not averaged immediately by opinion propagation. This increases the self-opinion of the leader who propagates this good opinion with a strong influence. Moreover, the leader propagates her good opinion about the agents that have a good opinion of her, which reinforces these agents in the population and their propagation of their good opinion about the leader. In summary, when an agent reaches some level of reputation (with all the others having very low reputations) she tends to lead the opinions, reinforcing herself and those who propagate a good opinion of her.
- When this leader's reputation becomes close to 1, it cannot grow anymore. Then a struggle with her potential rivals takes place. Indeed, in general, her allies, with the fluctuations, end up by having a lower opinion of the leader than the leader's self-opinion. In this case, the leader begins to decrease her opinion about this challenger. The leader is followed by the rest of the population with limited damage for her reputation if the challenger is not too strong. On the contrary, when the challenger has reached a reputation which is as strong as the one of the leader, then the fight is fatal for both of them and leads to come back to the generalised distrust.

- The rival that reaches the level of the leader and causes her loss can appear more or less rapidly. In some cases, the leader and her rival grow together right at the beginning and reach the top reputation almost at the same time. In this case, the dominance episode is short (first peak in the graph of dominance of pattern 4 around iteration 7000). Moreover, the rapid growth of the two leaders self-opinions leads them to decrease their opinions about all the rest of the population, by vanity. Figure 9 illustrates the rise and fall of a couple of leaders.
- On the contrary, when the leader manages to reach the top reputation with a significant difference with her followers, she may keep her dominance for more than 10 thousand time steps, with several attempts from challengers that she manages to stop soon enough. In this dominance episodes, there are periods where a progressive hierarchy is established with several agents having intermediate reputations. Figure 10 illustrates the rise of a single leader and figure 11 her successful struggle with a challenger.

The mechanisms behind pattern 3 are very similar, except that the leadership dominance are much longer, and that in the episodes of distrust, a hierarchical structure still persists. These explanations remain qualitative and for some aspects, conjectural. A more systematic and analytical study of these processes should be carried out in order to confirm and probably refine them.

5 Discussion - conclusion

Our initial intention was to explore the impact of Hobbes hypothesis about the role of vanity in human relations. In this respect, our work can be seen as using computer simulation as a tool to extend philosophical reflection: it allows one to make thought experiments with idealised populations of interacting agents. It is striking that our model leads to two situations that are the theoretical extremes that structure Hobbes theses: the general distrust where men are all enemies of each other (and of themselves), and the absolute dominance of one agent (the Leviathan). Moreover, when the absolute dominance of one agent takes place there are moments where a hierarchy of agents emerges because of the support of the leader, which seems to be to some extent in accordance with Hobbes thesis that the absolute dominance is a solution to the generalised distrust. It is even more striking that these two situations alternate dynamically in pattern 4, as if they were the two sides of the same coin.

Note that the model can also be related with other theories of social interactions, in particular the ones of René Girard ([Girard, 1972, 1982]). Indeed, with his thesis of mimetism as being the main driver of social interactions, Girard puts forward a mixture of the ingredients of our model: imitation (opinion propagation) and rivalry (vanity). Girard considers that these ingredients can lead to a state of generalised crisis, where all hierarchies are abolished, to the absolute dominance of one agent. Moreover, he insists on the intrinsic instability of the leadership relating it to the book of Job in the Bible. Our model is in accordance with these general views. However, Girard assumes also that the same mechanisms lead to scape goat structures where an agent is universally dispised. We have not observed this last pattern with our model. If this is confirmed that the model cannot generate such a pattern, it would be an argument against R. Girard's view that the same mechanisms lead to absolute dominance or to scape goat structures.

These remarks show that this model could help revisit some philosophical debates. The question is then: can it be related to actual social observations?

One could answer positively to this question, arguing that the patterns emerging from the model show major common features with familiar real social situations. For instance:

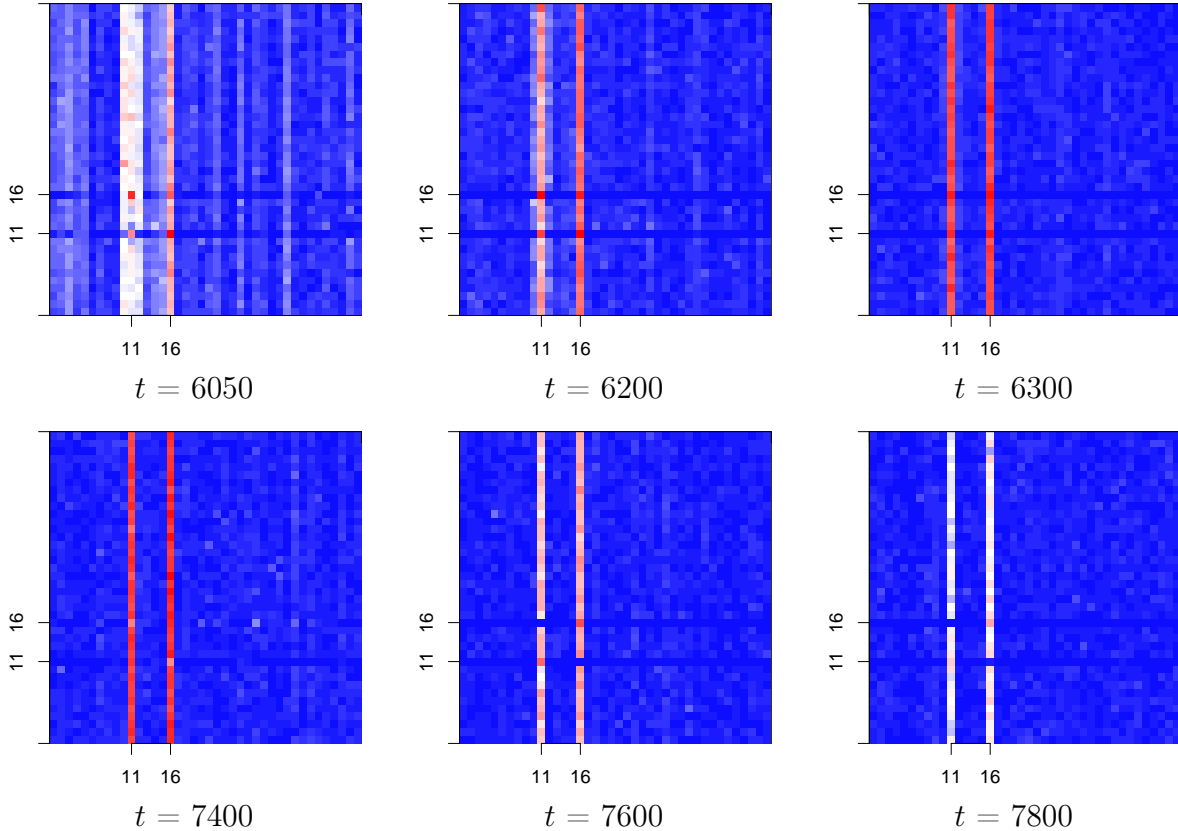


Figure 9: Pattern 4: Rise (first line of matrices) and fall (second line of matrices) of double leadership. At $t = 6050$, we observe that two agents ($i = 11$ and $i = 16$) support strongly each other (their reciprocal opinions are represented by the two red squares symmetric with respect to the diagonal), and both have positive self opinions. Moreover both tend to have very negative opinions of the rest of the population. At $t = 6200, 6300$, we observe that this process reinforces the reputations of both agents and decrease all the other opinions that become close to -1. At $t = 7400$, we observe that the two leaders do not support each other anymore, on the contrary, their opinion about each other (lighter squares) are lower than their respective reputations. Hence they enter in a negative loop of vanity in which they lead the rest of the population ($t = 7600, 7800$). The generalised distrust will take place after a few hundred iterations.

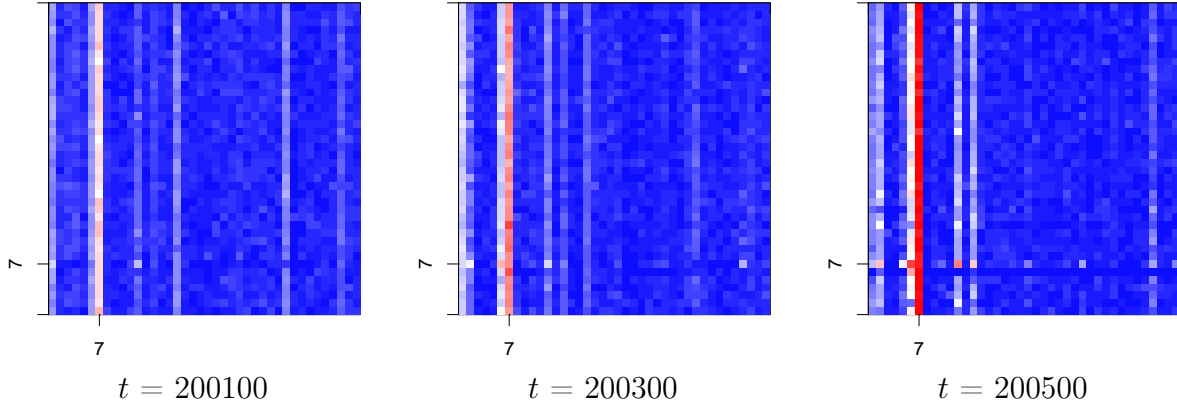


Figure 10: Pattern 4: Rising of a single leader. At $t = 200100$, we see that an agent (column with light pink cells at $i = 7$) has a higher reputation than the others, and she supports 3 other agents more than the average (white squares at $j = 1, 12, 17$). At $t = 200300$; we observe that her reputation significantly increased, that she supports other agents (particularly the pink square at $j = 6$). At $t = 200500$, the dominance of agent 7 is complete, and she supports strongly several agents that have their reputation rising (columns with several white or pink squares).

- Pattern 1 could be related to the tendency to privilege privatized lives in modern societies, turning exclusively to the small circle of family and close friends. Pattern 2 shows the risk of being excluded from this circle: the individual becomes isolated and loses her self esteem.
- Pattern 3 and pattern 4 would correspond to much more extreme situations, reminding for instance an absolute totalitarianism, where everybody is afraid of everybody (even parents fear their children) while all worship the absolute leader.
- The analysis of the patterns shows a positive bias for the agent self-opinion that is responsible for a global tendency for negative opinions about the others. This positive self bias is generally observed in social psychology [Hoorens, 1993]. The studies of links between self-esteem, leadership and groups (see e.g. Hogg and van Knippenberg [2003]) could be related to the mechanisms that we analysed in our model.
- In pattern 4, we observe that the leadership is given to individuals who tend to have a more positive view of the others than average and who are able to resist to the general opinion. This corresponds to the charismatic leader as defined in [van Knippenberg et al., 2004]. In our model, the dominant agents are able to make the others change strongly their opinions, which also corresponds to a recognised feature of leadership in the literature [Mary and Uhl-Bien, 2006; Hogg and van Knippenberg, 2003].

However, a closer analysis would lead to remain careful in these interpretations:

- The usual rule "the friends of my friends tend to be my friends" does not emerge from the dynamics of the model whereas it is generally found in observed social networks. This is surprising because actually the opinion influence should tend to implement this effect, by talking about friends. Therefore, we should check if for other values of the parameters, more realistic networks of friendship could emerge from the dynamics.
- Experiments show that a low self-esteem is very rarely expressed by subjects, whereas the majority of agents have a low self-esteem in patterns 3 and 4. A possible interpretation is that the usual subjects submitted to experiments are more likely to be closer to pattern 1 where everybody has a high self esteem. Moreover, as underlined previously our model

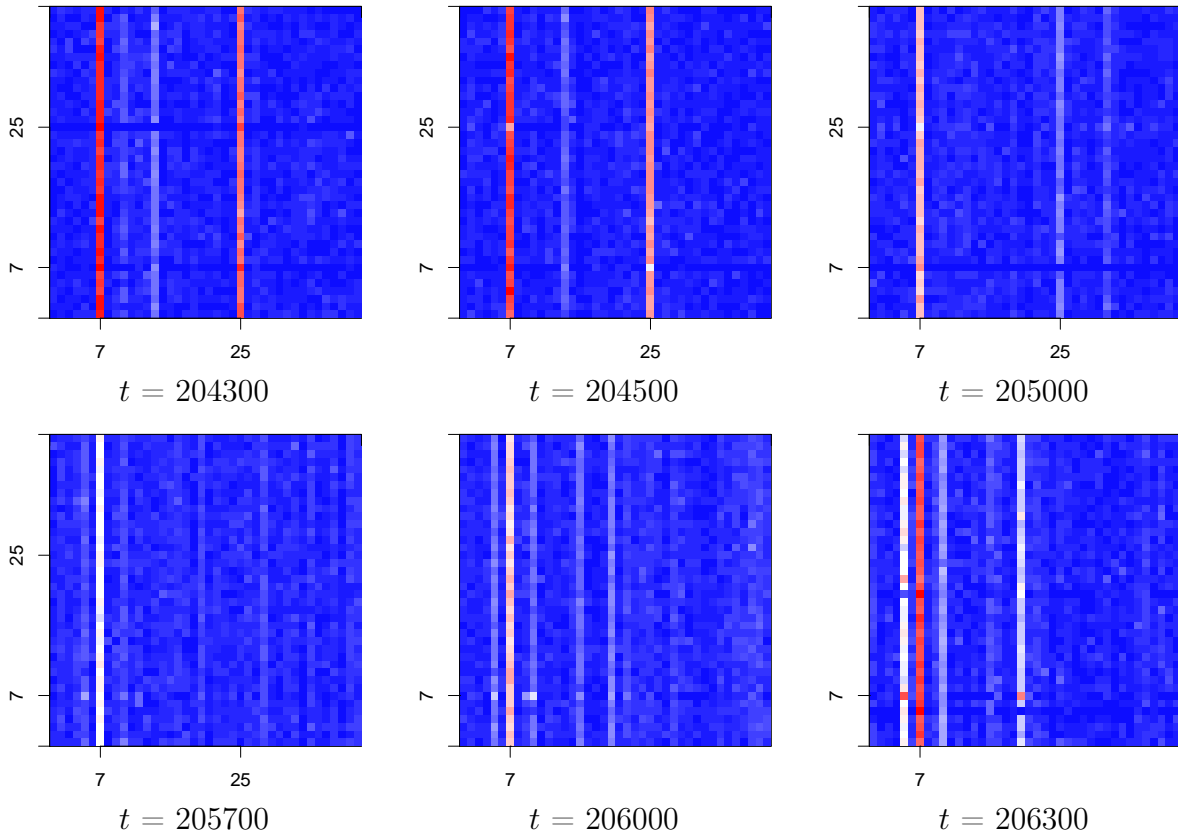


Figure 11: Pattern 4: the leader manages to get through the competition with a challenger. At $t = 204\,500$, the opinion of the challenger ($i = 25$) for the leader ($i = 7$) and of the challenger for the leader are significantly lower than their respective self opinions. They enter in a negative loop of vanity that ends up at $t = 205\,700$ with the challenger returning into the majority of very low reputations, and the leader having still some limited support from one other agent. The leader's reputation decreased much, but it is high enough to rise again ($t = 206\,000, 206\,300$) and a new episode of strong dominance will take place.

yields the usually observed positive self bias, compared with the other's opinions, hence this failure to match psycho-sociology observations about self-esteem should probably be discussed further.

- A major problem is that interpretations considering large societies are certainly too demanding for the model. In its present version, it is indeed limited to small populations, because it is supposed that everybody can meet with everybody and know everybody. Therefore, it is closer to a group dynamics.

In order to extend the study to larger sets of agents (thousands to millions for instance), it will be necessary to make new assumptions: to limit the number of agents that an agent can have in mind and discuss about, to limit the agents it can discuss with through different assumptions of a priori networks of interactions (possibly evolving with the opinions). Moreover, in the present setting, we considered identical agents at the beginning. It could be interesting to introduce some diversity, for instance in the propagation coefficient.

6 Acknowledgment

The work of T. Carletti has been partially supported by the FNRS grant "Mission Scientifique 2010-2011" allowing him to visit the Irstea - LISC during the fall 2010.

References

- Franco Bagnoli, Timoteo Carletti, Duccio Fanelli, Alessio Guarino, and Andrea Guazzini. Dynamical affinity in opinion dynamics modeling. *Phys. Rev. E*, 76:066105, Dec 2007. doi: 10.1103/PhysRevE.76.066105. URL <http://link.aps.org/doi/10.1103/PhysRevE.76.066105>.
- Katherine E. Buckley, Rachel E. Winkel, and Mark R. Leary. Reactions to acceptance and rejection: Effects of level and sequence of relational evaluation. *Journal of Experimental Social Psychology*, 40(1):14 – 28, 2004. ISSN 0022-1031. doi: 10.1016/S0022-1031(03)00064-7. URL <http://www.sciencedirect.com/science/article/pii/S0022103103000647>.
- Timoteo Carletti, Duccio Fanelli, and Righi Simone. Emerging structures in social networks guided by opinions' exchanges. *ACS*, 14:28, 2011. doi: 10.1142/S021952591100286X. URL <http://www.worldscinet.com/acs/14/1401/S021952591100286X.html>.
- C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009.
- G Deffuant. Comparing extremism propagation patterns in continuous opinion models. *Journal of Artificial Societies and Social Simulation*, 9(3):1–24, 2006. URL <http://jasss.soc.surrey.ac.uk/9/3/8.html>.
- G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(4):87–98, 2000.
- Gargiulo F. and S. Huet. Opinion dynamics in a group-based society . *EPL*, 91(58004):2–6, 2010.
- Steven J. Fein, Steven;Spencer. Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73(1):31–44, 1997. doi: 10.1037/0022-3514.73.1.31.

- S. Fortunato. Universality of the Threshold for Complete Consensus for the Opinion Dynamics of Deffuant et al. *International Journal of Modern Physics C*, 15:1301–1307, 2004.
- S. Galam. Sociophysics: a Review of Galam Models. *International Journal of Modern Physics C*, 19:409–440, 2008. doi: 10.1142/S0129183108012297.
- René Girard. *La violence et le sacré*. Grasset, 1972.
- René Girard. *Le bouc émissaire*. Grasset, 1982.
- Michael A Hogg and Daan van Knippenberg. Social identity and leadership processes in groups. volume 35 of *Advances in Experimental Social Psychology*, pages 1 – 52. Academic Press, 2003. doi: 10.1016/S0065-2601(03)01001-3. URL <http://www.sciencedirect.com/science/article/pii/S0065260103010013>.
- Vera Hoorens. Self-enhancement and superiority biases in social comparison. *European Review of Social Psychology*, 4(1):113–139, 1993. doi: 10.1080/14792779343000040. URL <http://www.tandfonline.com/doi/abs/10.1080/14792779343000040>.
- S. Huet and G. Deffuant. Openness leads to opinion stability and narrowness to volatility. *Advances in Complex Systems*, 13(3):405–423, 2010. doi: <http://dx.doi.org/10.1142/10.1142/S0219525910002633>. URL <http://cemadoc.cemagref.fr/cemoa/PUB00029104>.
- S. Huet, G. Deffuant, and W. Jager. A Rejection Mechanism In 2d Bounded Confidence Provides More Conformity. *Advances in Complex Systems*, 11(4):529–549, 2008.
- Mark R. Leary, Jean M. Twenge, and Erin Quinlivan. Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review*, 10(2):111–132, 2006. doi: 10.1207/s15327957pspr1002_2. URL <http://psr.sagepub.com/content/10/2/111.abstract>.
- J. Lorenz. Continuous Opinion Dynamics Under Bounded Confidence: a Survey. *International Journal of Modern Physics C*, 18:1819–1838, 2007.
- Mary and Uhl-Bien. Relational leadership theory: Exploring the social processes of leadership and organizing. *The Leadership Quarterly*, 17(6):654 – 676, 2006. ISSN 1048-9843. doi: 10.1016/j.leaqua.2006.10.007. URL <http://www.sciencedirect.com/science/article/pii/S1048984306001135>. <ce:title>The Leadership Quarterly Yearly Review of Leadership</ce:title>.
- Sanjay Srivastava and Jennifer S. Beer. How self-evaluations relate to being liked by others: Integrating sociometer and attachment perspectives. *Journal of Personality and Social Psychology*, 89(6):966 – 977, 2005. ISSN 2005-16185-010. doi: 10.1037/0022-3514.89.6.966. URL <http://www.sciencedirect.com/science/article/pii/S1048984304000864>.
- Yannick Stephan and Christophe Maiano. On the social nature of global self-esteem: A replication study. *The Journal of Social Psychology*, 147(5):573–575, 2007. doi: 10.3200/SOCP.147.5.573-576. URL <http://www.tandfonline.com/doi/abs/10.3200/SOCP.147.5.573-576>.
- K Sznadj-Weron. Sznajd model and its applications. *Acta Physica Polonica B*, 36(8), 2005.
- D. Urbig, J. Lorenz, and H. Herzberg. Opinion dynamics: The effect of number of peers met at once. *Journal of Artificial Societies and Social Simulation*, 11(2)(3):4, 2008. URL <http://jasss.soc.surrey.ac.uk/9/3/8.html>.

Daan van Knippenberg, Barbara van Knippenberg, David De Cremer, and Michael A. Hogg. Leadership, self, and identity: A review and research agenda. *The Leadership Quarterly*, 15(6):825 – 856, 2004. ISSN 1048-9843. doi: 10.1016/j.leaqua.2004.09.002. URL <http://www.sciencedirect.com/science/article/pii/S1048984304000864>.

J. V. Wood and A. L. Forest. Seeking pleasure and avoiding pain in interpersonal relationships. In M. D. Alicke and C. Sedikides, editors, *Handbook of Self-Enhancement and Self-Protection*, pages 258–78. New York, London: The Guilford Press, 2011.