Graphical Analysis of Agent-Based Opinion Formation Models

Carlos Andres Devia*[,1], Giulia Giordano[1,2]

April 2023

Abstract

Agent-based models of opinion formation in large scale populations are becoming increasingly complex, in part due to the embedding of individual psychological traits in the agents. Although the addition of these traits makes the models more realistic, it also results in models that are challenging to study from a theoretical perspective. In such cases, the characterisation of model properties relies mostly on simulation-based analyses. As the complexity and number of the models increases, so does the need for more techniques to analyse, contrast, and compare model properties. In this paper we present a novel graphical technique used to undercover behaviour patterns and capabilities of agent-based opinion formation models: the Agreement Plot. This technique can be used to characterise the relation between global opinion properties and initial opinions, agent parameters, and underlying digraphs. The proposed technique is applied to two opinion formation models: the classic Friedkin-Johnsen model, and the recently proposed Classification-based model.

Introduction

Agent-based models have been used in the study of opinion formation since the first works by French [1], Harary [2, 3], and DeGroot [4]. In an agent-based model, a collection of interconnected agents (representing individuals in a population) evolves according to internal dynamics and the information exchanged with other agents. This type of models have the advantage of being able to embed individual sociological and psychological opinion-formation mechanisms in the agent internal dynamics. Thanks to this capability, a wide range of models have been proposed in the literature, which include concepts from psychological and sociological research. Some of the features included in the models are *leaders* [5, 6], *emotions* [7, 8], *stubbornness* [9, 10], *biases* [11, 12, 13], *polarity* [14], *trust* [15, 16], *susceptibility* [17, 18], *mass media* [19], *tolerance* [20], *bounded confidence* [21], *weighted balance theory* [22], *controversy* [23], *assimilation* [24, 14], *coevolving networks* [25, 26], among others.

While the inclusion of these mechanisms makes the models more realistic, it also increases the difficulty of their analysis. As a result, the theoretical analysis of some of these models has to be restricted to special cases and the complete model capabilities and properties are left to be investigated from a numerical perspective. This tendency of exploring the model behaviour, properties, and capabilities from simulation-based analyses can be seen in the literature [27, 28, 29], and has resulted in a number of simulation-based techniques that use distributional measures to characterise the relation between initial conditions and the model outcomes. In [30] a histogram-based sorting algorithm is used to classify opinion distributions (sets of opinions) in real life and predicted by models in order to uncover opinion transitions seen in actual populations and predicted by different agent-based models. In [31] a systematic approach was taken to compute the probability that the final opinions have certain qualitative characteristics when only incomplete information on the initial conditions is available. In [14] the measures of Bias, Diversity, and Fragmentation are used to classify opinion distributions and relate model parameters with qualitative properties of the resulting opinions.

These type of simulation-based techniques form part of frameworks used to classify, compare, and contrast different opinion formation models [32]. The *Transition Tables* [30] and *Global Unifying Frame* [33] are examples of such frameworks.

In this paper we propose a novel analysis technique: the *Agreement Plot*. In this technique, every opinion distribution x is represented by two numbers: the mean $\overline{x}$ and the mean of the absolute values $\overline{|x|}$. Using these numbers as coordinates, it is possible to represent the opinion distribution as a point in the Cartesian plane. Although the information on the exact opinions is lost, the location of the opinion distribution in the Cartesian plane provides relevant qualitative and quantitative information on the opinion distribution. For instance, a point located near the origin represents societies where most agents have a neutral or indifferent opinion. A point with low mean and high absolute value mean corresponds to a society where agents have a strong opinion, but evenly distributed between agreement and disagreement, i.e. a polarised society.

*C.A.DeviaPinzon@tudelft.nl

[1]Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands

[2]Department of Industrial Engineering, University of Trento, 38123 Povo, Italy

The possibility of associating entire opinion distributions (with an arbitrary number of agents) with a single point in the Cartesian plane allows for the representation of multiple opinion evolutions in a single plot. When these opinion evolutions differ only by the initial opinions, agent parameters, or underlying digraph, these plots reveal how changes in these model parameters affect the final opinions. In addition to this, plotting predicted opinions for a wide array of initial opinions, agent parameters, and underlying digraphs provides a visual representation of the opinion distributions the models are able to produce. These type of analyses help understand how the model outcomes and behaviour is related to the model parameters, and also reveals intrinsic model properties.

The proposed technique can be applied to any agent-based opinion formation model in which the opinion every agent has belongs to a bounded interval in the real line. The techniques are exemplified in the classic Friedkin-Johnsen model [17, 18], and the recently proposed Classification-based model [34].

Methods

This section introduces the **Agreement Plot** concept. Based on this concept four different plots are defined: the Agent Parameter Time Evolution (**APTE**), Initial Opinion Time Evolution (**IOTE**), Agent Parameter Steady State (**APSS**), and Initial Opinion Steady State (**IOSS**) plots. Collectively, these four plots can be used to analyse models and identify behaviour patterns and intrinsic characteristics. The technique is exemplified with the Friedkin-Johnsen model.

Agreement Plot

Assume that a population has N individuals (or agents), and that each agent i has an opinion denoted x_i. The vector of all opinions $x = (x_i)_{i=1}^N$ is called an *opinion distribution*. Given an opinion distribution, the *general agreement* of x, denoted by $\pi(x)$ and defined as

$$\pi(x) = \left(\overline{|x|}, \overline{x}\right), \quad \text{where} \quad \overline{|x|} = \frac{1}{N}\sum_{i=1}^{N}|x_i| \qquad \text{and} \qquad \overline{x} = \frac{1}{N}\sum_{i=1}^{N}x_i, \tag{1}$$

represents how much the population cares about and agrees with a statement. Since $x_i \in [-1, 1]$ for all i, $\pi(x)$ is contained in the triangle with vertices in the points $(0, 0)$, $(1, -1)$, and $(1, 1)$. Therefore, the general agreement of every possible opinion distribution can be plotted in this triangle, independent on the number of agents. The plot of the general agreement of one or more opinion distributions in the Cartesian plane creates what we call the **Agreement Plot**.

The location of the point $\pi(x)$ provides useful information about the opinion distribution. For instance, if the point is located near one of the corners $(1, -1)$ or $(1, 1)$, the population cares significantly about that statement, as $\overline{|x|} \approx 1$ means most of the opinions are either -1 or 1 and therefore extremely strong opinions are common. Furthermore, near these corners $\overline{x} \approx 1$ or $\overline{x} \approx -1$, so almost every agent must have the same opinion, either 'strongly agree' or 'strongly disagree'.

If the point $\pi(x)$ is in the neighbourhood of the corner $(0, 0)$, then $\overline{|x|} \approx 0$, so most agents are indifferent about that particular statement, and the average opinion $\overline{x} \approx 0$ is neutral. If $\pi(x) \approx (1, 0)$, the agents care about that statement but the mean is low, therefore two opposite groups must exist, with almost equal number of agents, i.e. polarisation. When $\pi(x)$ is located near the lines $y = \pm x$, almost all the population either agrees or disagrees to some degree. On the other hand, if $\pi(x)$ is located near the line $y = 0$, there is almost equal amount of agreement and disagreement.

The possibility to represent any opinion distribution by a single point with simple and intuitive interpretation allows for new types of model analysis. Specifically, the Agreement Plot can be used in two distinct ways: *time evolution* (TE), and *steady state* (SS) analysis.

Colour coding: In order to maximise the information obtained by the different plots, the plotted lines and points can be colour coded. Unless otherwise specified, the colour coding will be model-dependent. The plot line and point colour represents the average agent parameters of a given society. For each of the considered models, the colour coding is:

- For the Friedkin-Johnsen model, the agent parameter is susceptibility (the susceptibility of agent i is denoted λ_i) and is a number in the interval $[0, 1]$. The point or line colour represents the mean population susceptibility $\overline{\lambda}$; less susceptible populations have a dark magenta colour, and more susceptible populations have a teal colour. This colour coding can also be seen in the agent parameter histograms from Fig 4.

- For the Classification-based model, each agent has three parameters: the weight of the conformist, radical, and stubborn trait (denoted α_i, β_i, and γ_i, respectively, for each weight, for agent i). The three weights are in the interval $[0, 1]$ and they add 1. The corresponding point or line colour is the RGB colour obtained

by combining $\overline{\alpha}$ blue, $\overline{\beta}$ red, and $\overline{\gamma}$ green; where $\overline{\alpha}$, $\overline{\beta}$, and $\overline{\gamma}$ are the corresponding average conformist, radical, and stubborn trait weights. This colour coding can be seen in the agent parameter ternary diagram from Fig 9 (an explanation on how to interpret the ternary diagrams can be found in the Supplementary Information).

When the model is clear from context, the term 'agent parameter' is interchangeable with the agent parameter name. For instance, the statements 'higher parameter mean' is the same as 'higher susceptibility mean', if from the context it is clear that the statements refer to the Friedkin-Johnsen model.

Agreement Plot of Time Evolution

We first use the Agreement Plot to plot one or more parametric curves (with the parameter being time) corresponding to the temporal evolution of one or more different populations.

Fig 1 shows an example of the complete evolution of a population's opinion represented by a single parametric curve in the Agreement Plot. On the top left figure we can see the opinion evolution of every single agent over time. The bottom left figure shows the time-dependence of the mean of the opinion absolute values and of the opinion mean for the complete population. We can represent these two lines as a parametric curve in the Agreement Plot, as shown in the right plot. Visualising the time evolution of the individual opinions provides some information, however collective behaviours can be hidden in the complex evolution of all individual agents. The Agreement Plot allows us to better visualise collective behaviours.

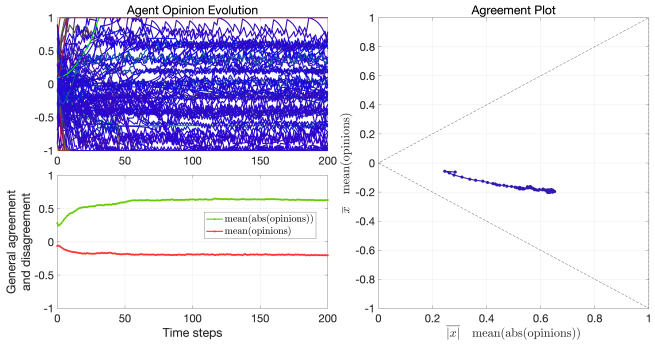


Figure 1: Example of how a parametric curve in the Agreement Plot can represent the opinion evolution of a complete population. The top left plot shows the time evolution of all the agents in a population. The bottom left plot shows the time-dependence of the mean of the opinion absolute values $\overline{|x|}$ and of the opinion mean $\overline{x}$ for the complete population. The right plot shows the corresponding parametric curve in the Agreement Plot.

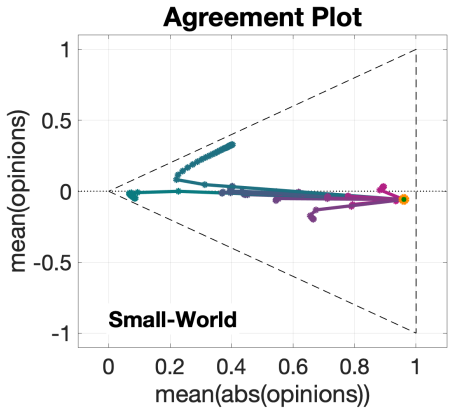
If the populations differ in only one aspect, for instance, agent parameters, or initial opinions, then the time evolution Agreement Plots show the overall effect these changes have on the opinion evolution. Fig 2 shows the two approaches applied to the Friedkin-Johnsen model.

Fig 2a shows 15 parametric curves. In these plots the initial opinion and underlying digraph are constant and only the agent parameters change (this is represented by the change in the line colour). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. It can be seen that lines with more dark magenta colour tend to be shorter than lines with more teal colour. This is to be expected, as for the first lines the agents are less susceptible and therefore the opinions change less. On the other hand, teal lines correspond to populations with high average susceptibility and therefore opinions change significantly more.

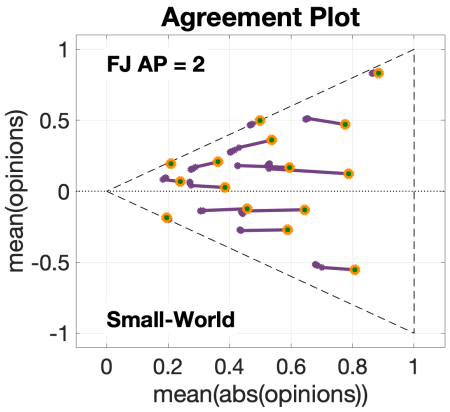
Fig 2b shows 15 parametric curves corresponding to opinions evolving with the same agent parameters and underlying digraph, and different initial opinions (represented by the orange dots). The agent parameter number can be seen in the upper left corner (Fig 4 shows a histogram of the corresponding agent parameters). From the figure, it is clear that independent of the initial opinions there is a tendency to move towards the left and opinions close to the lines $y = \pm x$ tend to remain in the same place.

Although plots in Fig 2 evidence certain behaviours, a valid question to ask is if these observations depend on the chosen initial opinions, agent parameters, and underlying digraphs. For instance, how would Fig 2a change if the digraph topology changed from Small-World to Complete? or to Scale-Free? These questions can be answered by repeating similar plots for a variety of different initial opinions, agent parameters, and underlying digraphs, as shown in Figs 5 and 6.

For simplicity, plots of parametric curves with different agent parameters and constant initial opinions and underlying digraph (like the one shown in Fig 2a) will be referred to as Agent Parameter Time Evolution (**APTE**) plots. And plots of parametric curves with different initial opinions, and constant agent parameters



(a) Parametric curves for 15 different opinion evolutions with the same initial opinions (orange dot) and underlying digraph, and different agent parameters. The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. This type of plot is called Agent Parameter Time Evolution (**APTE**).



(b) Parametric curves for 15 different opinion evolutions with the same agent parameters and underlying digraph, and different initial opinions. The agent parameter number can be seen in the upper left corner (Fig 4 shows a histogram of the corresponding agent parameters). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. This type of plot is called Initial Opinion Time Evolution (**IOTE**).

Figure 2: Use of the Agreement Plot to identify opinion evolution behaviours produced by the Friedkin-Johnsen model. In each of the panels, multiple parametric curves (where the parameter is time) are plotted that correspond to populations that differ only in the agent parameters (Fig 2a), or initial opinions (Fig 2b). For all the plots the line colour represents the average susceptibility of the agents (magenta is less susceptible, teal is more susceptible). The initial opinions are the orange dots, and when either the agent parameters are constant, their number is also shown in the plot. All simulations were for 50 time steps and 100 agents.

and underlying digraph (like the one shown in Fig 2b), will be referred to as Initial Opinion Time Evolution (**IOTE**) plots.

Agreement Plot of Steady State

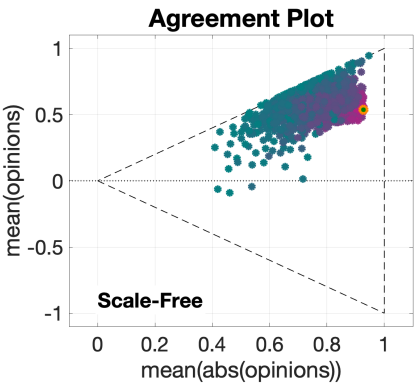
If we only care about the steady state, it is possible to skip the parametric curve and simply plot the initial and final opinion distributions. Doing this for a wide collection of agent parameters, and initial opinions (while keeping the other model parameters constant) can highlight what kind of opinion distributions the model can produce starting from a given point and also how the model transforms the opinions. For these plots, the number of time steps can also be increased, so that the models are given enough time to show the opinion changes caused by slow dynamics. For these plots, the number of time steps is increased from 50 to 1000.

We first consider the case of multiple agent parameters. These plots are straightforward: we simply repeat the **APTE** plots but without the parametric curve, and with (possibly) a higher number of agent parameters. Fig 3a shows an example for the Friedkin-Johnsen model.

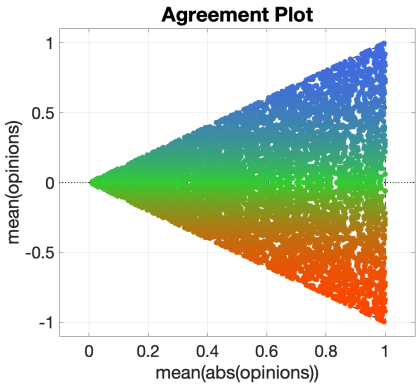
Fig 3a shows the Agreement Plot of 3528 predicted opinions starting with the same initial opinion distribution (orange dot), underlying digraph, and different agent parameters (as can be seen by the different points with dark magenta and teal colour combination). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. Besides observing behaviours consistent with previous simulations (such that the dark magenta points are closer to the orange dot, in contrast to the teal points), it is interesting to note that very few opinions can be found below the x-axis or to the right of the initial opinions. This seems to suggest that, independent of the agent parameters, the sign of the mean will most likely remain constant, and also that the mean of the opinion absolute values is unlikely to increase for this model.

Adapting **IOTE** plots to represent the opinion evolution of a wide variety of initial opinions for a fixed agent parameter set and underlying digraph cannot be done as with the **APSS** plots, simply because there is not enough space. Instead what can be done is to start with a reference plot, like the one in Fig 3b, where each dot corresponds to a different initial opinion (and has a unique colour). Then, evolve the model, with the same agent parameters and underlying digraph, for each of the initial opinions, and present the resulting opinions in a separate plot, as shown in Fig 3c. Two points, one from Fig 3b and other from Fig 3c with the same colour represent the same population, the former is the initial opinion distribution, and the later, the final opinion distribution.

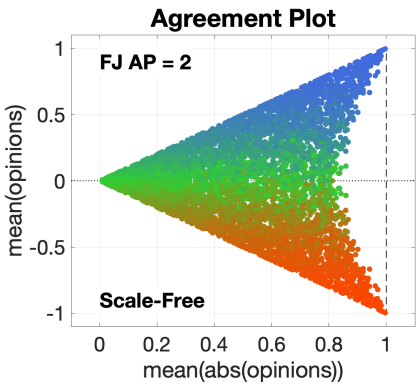
Fig 3c shows how the 5314 initial opinions from Fig 3b transform according to the Friedkin-Johnsen model when the agent parameters and underlying digraph are the ones in the top and bottom left corners of Fig 3c,



(a) Agreement Plot of the predicted opinions for 3528 different opinion evolutions with the same initial opinions (orange dot) and underlying digraph, and different agent parameters. The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. This type of plot is called Agent Parameter Steady State (**APSS**).



(b) Reference image for the location in the Agreement Plot of the 5314 initial opinions used in the Initial Opinion Steady State (**IOSS**) plot.



(c) Agreement Plot of the predicted opinions for 5314 different opinion evolutions with the same agent parameters, underlying digraph, and different initial opinions, shown in Fig 3b. The agent parameter number can be seen in the upper left corner (Fig 4 shows a histogram of the corresponding agent parameters). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. This type of plot is called Initial Opinion Steady State (**IOSS**).

Figure 3: Use of the Agreement Plot to display the capacity of the Friedkin-Johnsen model to create a wide range of opinion distributions. For the plot in Fig 3a the point colour represents the average susceptibility of the agents (magenta is less susceptible, teal is more susceptible), and the initial opinion point is orange. Dots with the same colour in Figs 3b and 3c correspond to initial and final opinion distributions for the same simulation (in these plots the colours don't have a meaning on themselves, they are only used to visually connect one dot in Fig 3b with a dot in Fig 3c). Fig 3b shows the initial opinion, whereas Fig 3c shows the predicted opinions. All simulations were for 1000 time steps and 100 agents.

respectively. Several observations can be made about Fig 3c. The contraction of the initial opinions towards the y-axis is evident, and also the fact that the contraction is more pronounced near the x-axis. This indicates that very polarised opinions are transformed into less polarised opinions, which is consistent with the averaging tendency caused by the susceptibility trait in the Friedkin-Johnsen model. The colours also allow to see that the mean opinion remains almost constant, this is hinted by the colour gradient seen in Fig 3b, which is still distinguishable in Fig 3c.

As with **APTE**, and **IOTE** plots, the analysis possibilities are increased when multiple plots of the same type are arranged together, as shown in Figs 7 and 8. This allows us to determine if the observed behaviour is intrinsic for the model, or if (and how) it depends on the model parameters. It is worth clarifying that, for all **IOSS** plots, Fig 3b will always be the reference for the starting initial opinions.

The information provided by, and possible interpretations of, the four previously introduced plots is summarised in Table 1.

Plot name	What is plotted	Constant model parameters	Variable model parameters	Example figure
APTE	Parametric curves (the parameter is time) corresponding to the Agreement Plot of the system evolution (50 time steps).	Initial opinions and underlying digraph	Agent parameters	Fig 2a
IOTE		Agent parameters and underlying digraph	Initial opinions	Fig 2b
APSS	Agreement Plot of the steady state opinions resulting from the system evolution (1000 time steps).	Initial opinions and underlying digraph	Agent parameters	Fig 3a
IOSS		Agent parameters and underlying digraph	Initial opinions	Fig 3c

Table 1: Summary of the four plot types that make use of the Agreement Plot to analyse agent-based opinion formation models.

Digraph	APL	D	$\overline{CC}$	$\sigma(CC)$	#E	$\overline{\delta^{out}}$	$\sigma(\delta^{out})$	$\overline{\delta^{in}}$	$\sigma(\delta^{in})$	BC	+/-
Directed Ring	6.0704	12	0.26477	0.00642	634	6.34	2.3479	6.34	1.661	0	1.3605
Scale-Free	2.9297	7	0.77612	0.17639	201	2.01	1.6464	2.01	1.7272	0.47525	2.0722
Small-World	2.6232	5	0.08205	0.0043965	774	7.74	3.164	7.74	5.8307	0.33828	1.171
Complete	1.5099	2	0.48936	0.00010153	4952	49.52	22.6158	49.52	22.8582	0.50041	0.98098

Table 2: Metrics for the digraph topologies considered in the Agreement Plot analysis. The metrics are Average Path Length (APL), Diameter (D), Mean clustering ($\overline{CC}$), Clustering variance ($\sigma(CC)$), Number of edges (#E), Mean out-degree ($\overline{\delta^{out}}$), Out-degree variance ($\sigma(\delta^{out})$), Mean in-degree ($\overline{\delta^{in}}$), In-degree variance ($\sigma(\delta^{in})$), Bidirectional Coefficient (BC), and Positive to negative edge ratio (+/-) (applies only for the Classification-based model). An explanation on how these metrics are computed is provided in the Supplementary Information.

Results

We can apply the proposed methodology to the classical Friedkin-Johnsen [17, 18] model and the recently proposed Classification-based [34] model. In both cases the underlying digraphs have one of four possible topologies: Directed Ring, Scale-Free, Small-World, and Complete. Some metrics of these topologies are shown in Table 2.

Friedkin-Johnsen Model Analysis

Fig 4 shows the histograms corresponding to the agent parameters used for the **APTE**, **APSS**, **IOTE**, and **IOSS** plots for the Friedkin-Johnsen model.

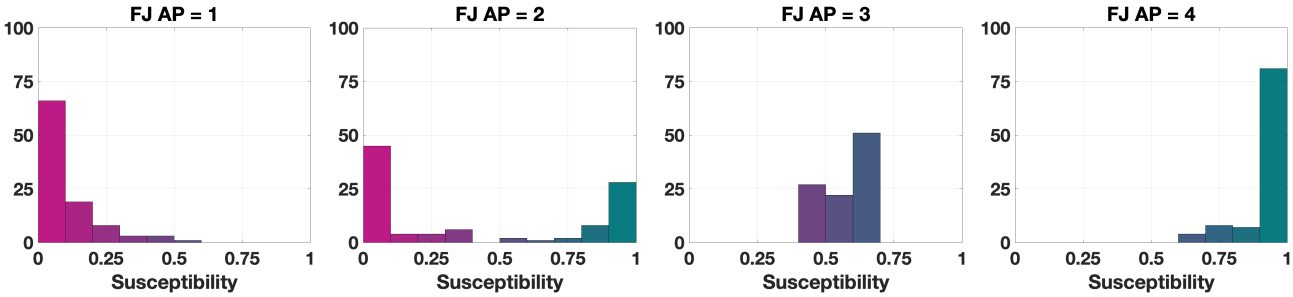


Figure 4: Histograms of the agent parameters used in the **APTE**, **APSS**, **IOTE**, and **IOSS** plots for the Friedkin-Johnsen model (Figs 5, to 8 respectively).

Fig 5 shows 12 different **APTE** plots for 3 choices of the initial opinions and 4 different underlying digraphs. The metrics for these four digraphs are summarised in Table 2.

Comparing the plots in Fig 5, it is possible to conclude that, although the overall qualitative behaviour is the same as seen in Fig 2a, it is clear that the network topology has an important effect on the opinion evolution. Specifically, for all the initial opinions it can be noted that the opinions change less for the Scale-Free digraph, whereas for the other three digraphs there appears to be no significant difference (this is particularly true for row 2). A possible explanation for this phenomenon is that the high network diameter and clustering variance, combined, lead to a network that has a few central, highly influential vertices connecting otherwise disconnected subnetworks. Because of this, opinions take more time to spread, possibly explaining why opinions change less than with the other digraphs for all initial opinions.

It is also interesting to remark that the other three topologies are quite different, and still, simulation results are very similar. This is specially true for the Ring and Complete digraphs (columns 1 and 4), even though, as seen in Table 2, for some metrics one has the highest value, while the other has the lowest value (for instance average path length, or diameter). This suggests that, for the Friedkin-Johnsen model, the opinion evolution is independent of these network metrics.

Fig 6 shows 12 **IOTE** plots for 3 choices of the agent parameters and 4 different underlying digraphs. These plots provide further evidence that the observations made from Fig 5 are true for opinions starting from any initial opinion distribution. This is important to check, as it may not always be true. It is also interesting to note that the parametric curves for the plot in row 3, column 2 (Scale-Free digraph with highly susceptible agents) show a peculiar behaviour in that they don't immediately move to the lines $y = \pm x$, like the other plots with these same agent parameters do. This can possibly be explained by the structure of the Scale-Free network. In contrast to the other digraphs which are made of a single 'entity', the 'different compartments' of the Scale-Free network disperse and slow down the tendency of the Friedkin-Johnsen model (with high agent susceptibility) to create perfect consensus.

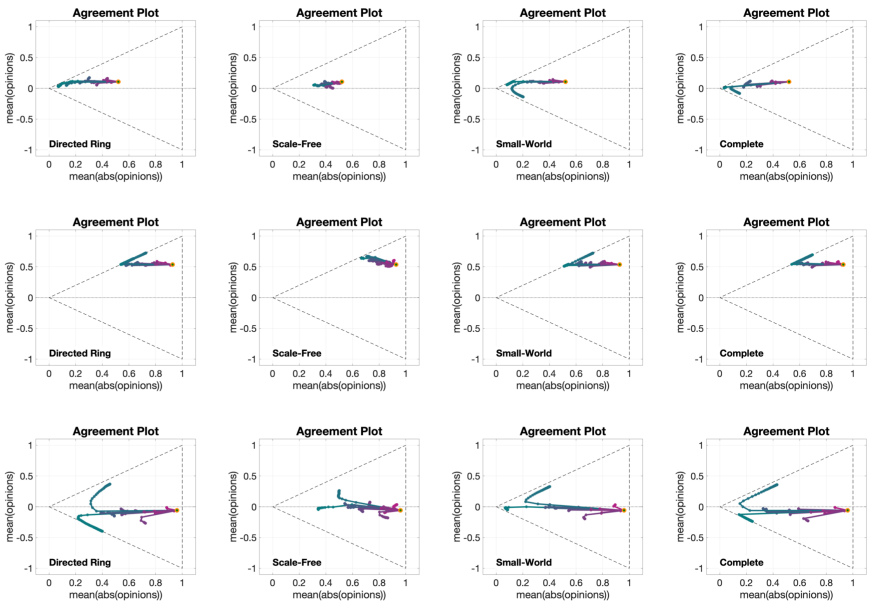


Figure 5: **APTE** plots for the Friedkin-Johnsen model. Each of the 12 **APTE** plots includes 15 curves with different choices of agent parameters and constant initial opinion distributions and underlying digraphs. Plots in the same row start from the same initial opinion distribution (represented by the orange dot). Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete).The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 50 time steps.

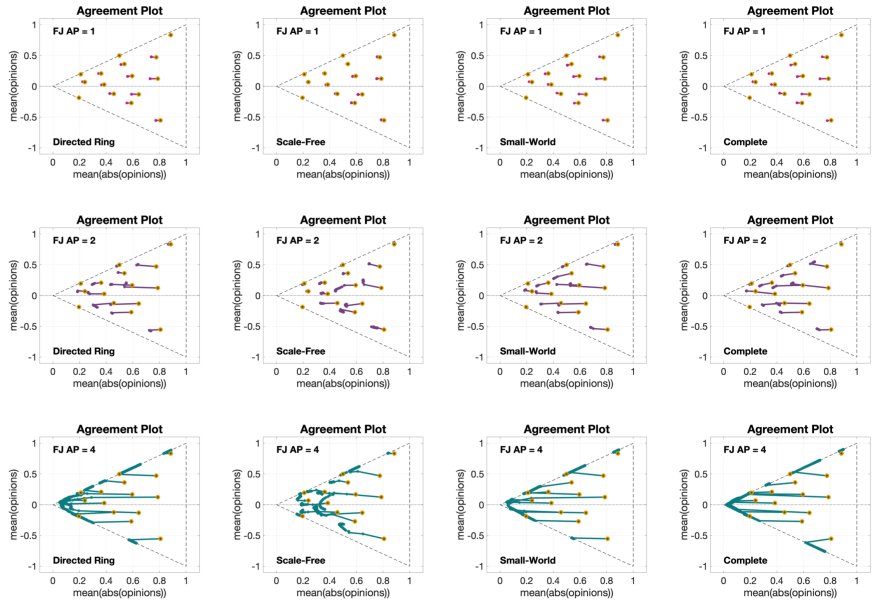


Figure 6: **IOTE** plots for the Friedkin-Johnsen model. Each of the 12 **IOTE** plots includes 15 curves with different choices of initial opinion distributions and constant agent parameters and underlying digraphs. Plots in the same row have the same agent parameters. Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete).The agent parameter number can be seen in the upper left corner (Fig 4 shows a histogram of the corresponding agent parameters). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 50 time steps.

The **APSS** 12 plots, for 3 different initial opinions, and 4 different underlying digraphs in Fig 7 indicate that, for all initial opinions and digraphs, the qualitative effect of increasing the mean susceptibility is the same: as the mean susceptibility increases, the dark magenta points form a 'cone' to the left of the initial point. The higher the mean susceptibility, the more the predicted opinions move to the left. This trend continues until they arrive to the $y = \pm x$ lines, at which point they move along these lines, which is where most of the teal

points can be found.

This also implies that the higher the mean susceptibility, the more likely the final opinions are to form perfect consensus, which is not surprising. What is surprising, is the noticeable difference between the second column and the others. As in Fig 5 we see the Scale-Free digraph effect in that opinions don't change as much as with the other topologies. This is better evidenced in the plots of row 3.

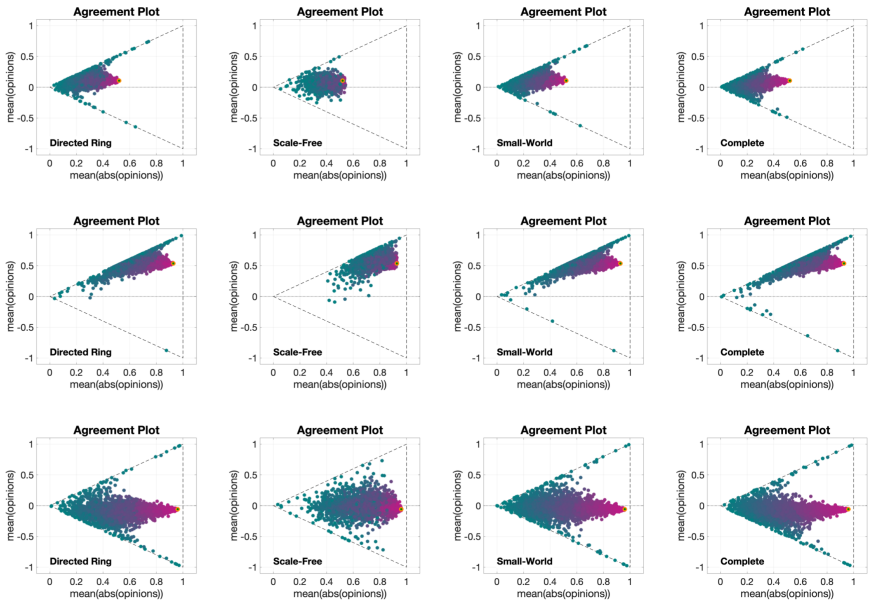


Figure 7: **APSS** plots for the Friedkin-Johnsen model. Each of the 12 **APSS** plots includes 3528 points with different choices of agent parameters and constant initial opinion distributions and underlying digraphs. Plots in the same row start from the same initial opinion distribution (represented by the orange dot). Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete).The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 1000 time steps.

Finally, Fig 8 presents 12 **IOSS** plots which show how the reference initial opinions seen in Fig 3b evolve for 3 different choices of the agent parameters and 4 different underlying digraphs. These plots showcase familiar behaviours, while at the same time providing additional information on the model. It is not surprising that every plot contracts, as seen initially in Fig 3c: every plot so far shows this pattern. Nevertheless, comparing the plots evidences a hidden nonlinearity in the 'contraction factor'. Indeed, the agent parameters in row 2 are in a way the middle point between the agent parameters in row 1 and 3, however, the difference between plots in rows 1 and 2 is less significant than the difference between plots in rows 2 and 3. Regarding the underlying digraph, row 1 shows that for models with very low mean susceptibility the result is almost independent of the digraph topology. The opposite can be seen in rows 2 and 3, where the Scale-Free digraph again causes a significantly different behaviour.

After analysing Figs 5 to 8, we can draw the following conclusions regarding the opinion transformation and predictive capabilities of the Friedkin-Johnsen model:

- Independent on the initial opinions, agent parameters, and underlying digraph the opinions have the initial tendency to move towards the lines $y = \pm x$, where all the agents either agree or disagree.
- When the mean susceptibility is below about 0.5, or when the underlying digraph has a Scale-Free topology, the opinions change more slowly and the final asymptotic value is closer to the initial value.
- As the mean susceptibility increases, or the average path length and diameter of the underlying digraph decreases, the opinions change faster and their final asymptotic value is closer to the line $y = \pm x$.
- When the mean susceptibility is high, or the average path length and diameter of the underlying digraph are low, the opinions reach the $y = \pm x$ line and move parallel to it.
- The final predicted opinions are very unlikely to have a higher mean of the opinion absolute values than the initial opinions, and, when that happens, either all agents agree or disagree.
- In a society with low susceptibility variance, opinions will change more than another society with similar susceptibility mean and higher susceptibility variance.

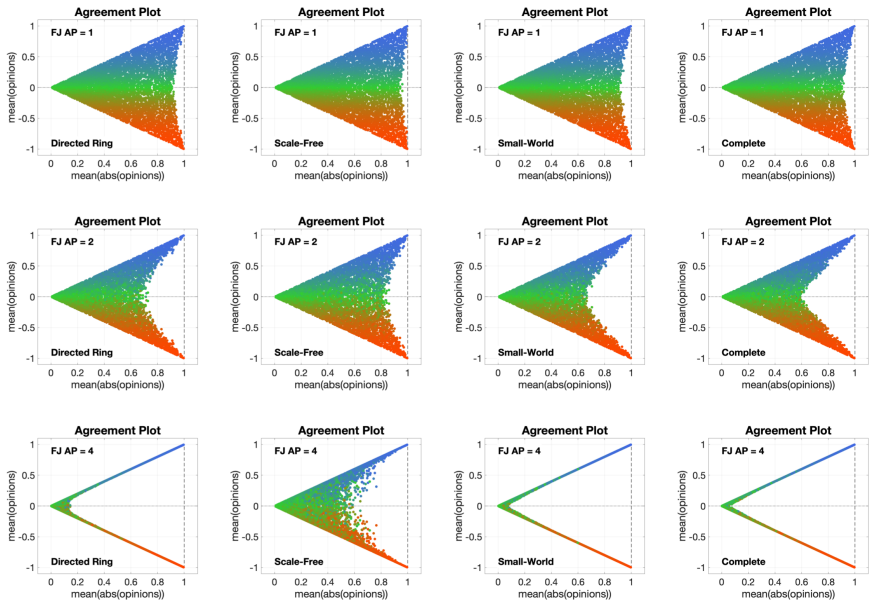


Figure 8: **IOSS** plots for the Friedkin-Johnsen model. Each of the 12 **IOSS** plots includes 5314 points with different choices of initial opinion distributions and constant agent parameters and underlying digraphs. Plots in the same row have the same agent parameters. Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete).The agent parameter number can be seen in the upper left corner (Fig 4 shows a histogram of the corresponding agent parameters). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 1000 time steps.

Classification-based Model Analysis

Figs 10 to 13 present multiple plots analogous to Figs 5 to 8. For Figs 10, 11, 12, and 13, the underlying digraphs have the same topology as the digraphs used in the Agreement Plot analysis of the Friedkin-Johnsen model. Therefore, the topology metrics are shown in Table 2.

Fig 9 shows the agent parameters used for the **APTE**, **APSS**, **IOTE**, and **IOSS** plots for the Classification-based model. Since for this model the agent parameters are three weights between 0 and 1, that add 1, they can be naturally represented in a ternary diagram (an explanation on how to interpret the ternary diagrams can be found in the Supplementary Information).

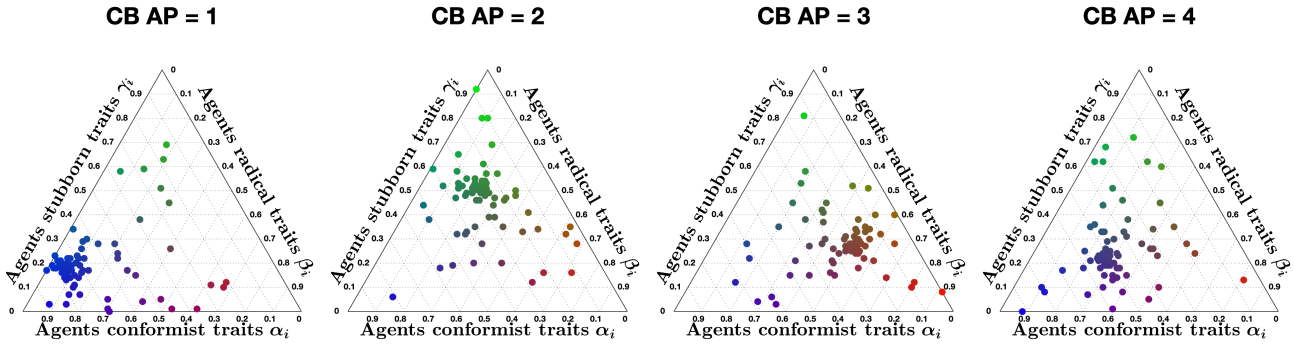


Figure 9: Ternary plots of the agent parameters used in the **APTE**, **APSS**, **IOTE**, and **IOSS** plots for the Classification-based model (Figs 10, to 13 respectively).

We will associate with each trait a colour, for conformism ($\overline{\alpha}$) blue, for radicalism ($\overline{\beta}$) red, and for stubbornness ($\overline{\gamma}$) green. Thus the line or point colour associated with a given society is the colour resulting from mixing the amount of blue, red, and green corresponding to the average weights in that particular society. So, for instance, a teal colour may indicate an equal amount of conformism and stubbornness and very low radicalism, while a red line indicates a highly radical society or a blue line a high level of conformism.

Fig 10 shows 12 **APTE** plots for 3 different initial opinions and 4 different underlying digraphs. These **APTE** plots show that the parametric curves not only move towards the y-axis but also move away from it. There are also some curves that appear to move towards the centre of the Agreement Plot, but this behaviour

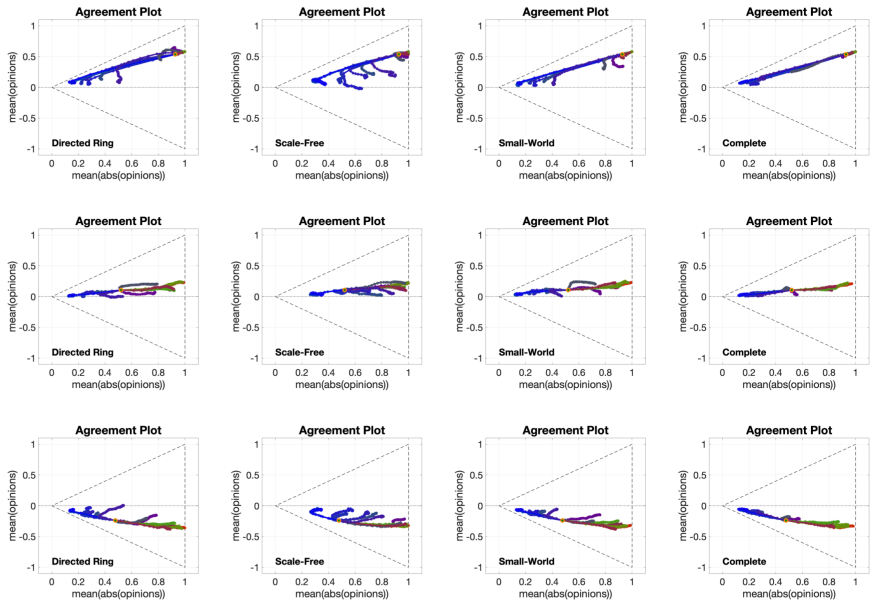


Figure 10: **APTE** plots for the Classification-based model. Each of the 12 **APTE** plots includes 15 curves with different choices of agent parameters and constant initial opinion distributions and underlying digraphs. Plots in the same row start from the same initial opinion distribution (represented by the orange dot). Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete).The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 50 time steps.

can only be observed for some initial opinions. Looking at the line colours it is clear that the lines moving towards the y-axis are mainly blue, representing highly conformist societies. This makes sense, because in the Classification-based model the conformist trait produces consensus, and thus it is expected that the blue lines behave like the Friedkin-Johnsen lines.

On the other hand, the curves moving away from the y-axis are mostly red and green. For the red lines radicalism tends to move the agent opinions to the extremes, resulting in a higher mean of the absolute values. For the green lines the behaviour is unexpected, however upon closer inspection it is clear that these green lines also have a significant red component and this mild radical trait may be responsible for this change.

Regarding noticeable differences due to the underlying digraph, societies evolving over the Scale-Free digraph tend to move more towards the centre of the Agreement Plot, and in contrast the other societies (especially the ones with Complete digraph) move more to the right or left of the plot. One possible explanation could be that, since the Classification-based model allows for differences between the expressed opinion and how it is perceived by influenced neighbours, in networks with longer average path length these 'errors' accumulate, thus diluting the trait's effects and for instance preventing a highly conformist society from reaching consensus (blue curves not going completely to the $y = \pm x$ lines).

Fig 11 shows 12 **IOTE** plots for 3 different agent parameters and 4 different underlying digraphs. These plots suggest that, given a set of agent parameters and underlying digraph, all the initial opinions will move towards the same region in the Agreement Plot: this region just depends on the agent parameters and underlying digraph. This is a new behaviour. For the Friedkin-Johnsen model, all the curves moved to the same direction. In contrast, for the Classification-based model the general direction the curves follow depends on the initial opinion location in the Agreement Plot and the region to which all curves converge to.

The convergence speed (inferred from the distance between points in the parametric curves) seems to depend only on the agent parameters and not the underlying digraph. The distance between points in the green curves is noticeable less than in the other curves, indicating that societies with mostly stubborn traits change more slowly, which is coherent, given that the stubborn trait has the tendency of keeping the opinions unchanged, thus decreasing the opinion change caused by the other agents' traits.

Fig 12 shows 12 **APSS** plots for 3 different initial opinions and 4 underlying digraphs. These **APSS** plots look remarkably different from the same type of plots for the Friedkin-Johnsen model. The first difference is that, unlike all the other **APSS** plots, the plots in Fig 12 have few points located along the lines $y = \pm x$. The points mostly appear to converge to a connected (and in most cases convex) subset of the Agreement Plot. In some cases the initial point is clearly located in the interior of this region, while in others it is at the boundary. Unlike the Friedkin-Johnsen model, the range of opinions that can be achieved is not too wide. This is particularly true when looking at the y-axis components of the points: most predicted points in the

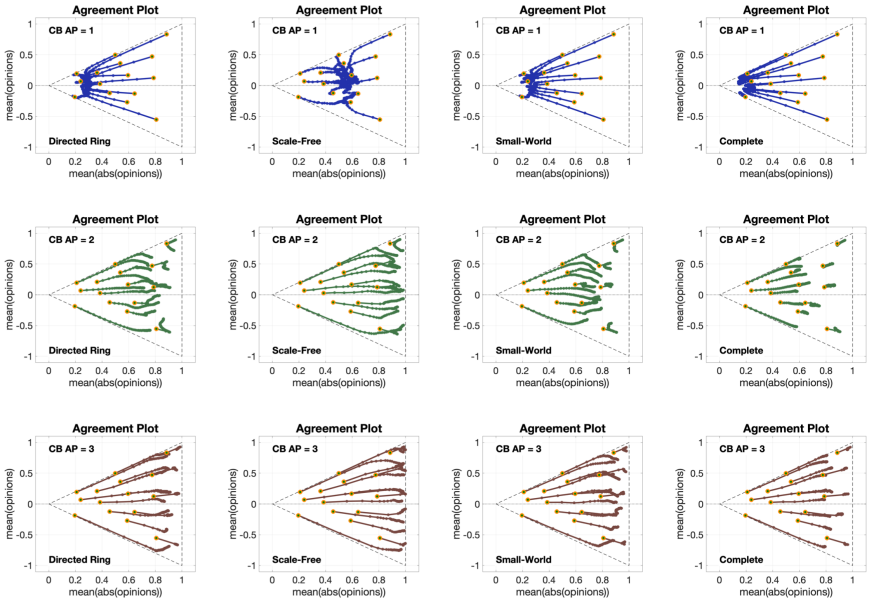


Figure 11: **IOTE** plots for the Classification-based model. Each of the 12 **IOTE** plots includes 15 curves with different choices of initial opinion distributions and constant agent parameters and underlying digraphs. Plots in the same row have the same agent parameters. Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete). The agent parameter number can be seen in the upper left corner (Fig 9 shows a representation of the corresponding agent parameters). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 50 time steps.

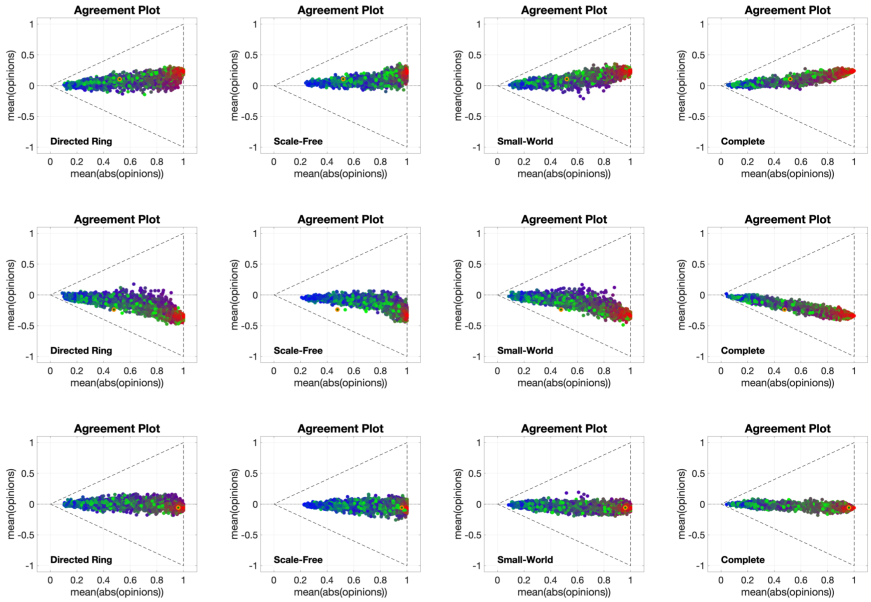


Figure 12: **APSS** plots for the Classification-based model. Each of the 12 **APSS** plots includes 3528 points with different choices of agent parameters and constant initial opinion distributions and underlying digraphs. Plots in the same row start from the same initial opinion distribution (represented by the orange dot). Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete).The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 1000 time steps.

Classification-based model have a y component not too far from the initial opinions. This means that the Classification-based model tends to leave the mean of the opinions relatively unchanged.

Looking at the dot colours, we can see the right dots are mostly red, the left dots are mostly blue, and the green dots can be found in the vicinity of the initial opinion distribution point in either direction. This indicates that mostly conformist societies tend to less extreme opinions, mostly radical societies tend to more extreme

opinions, and mostly stubborn societies don't change much and the change direction is determined by the trait with the second highest weight.

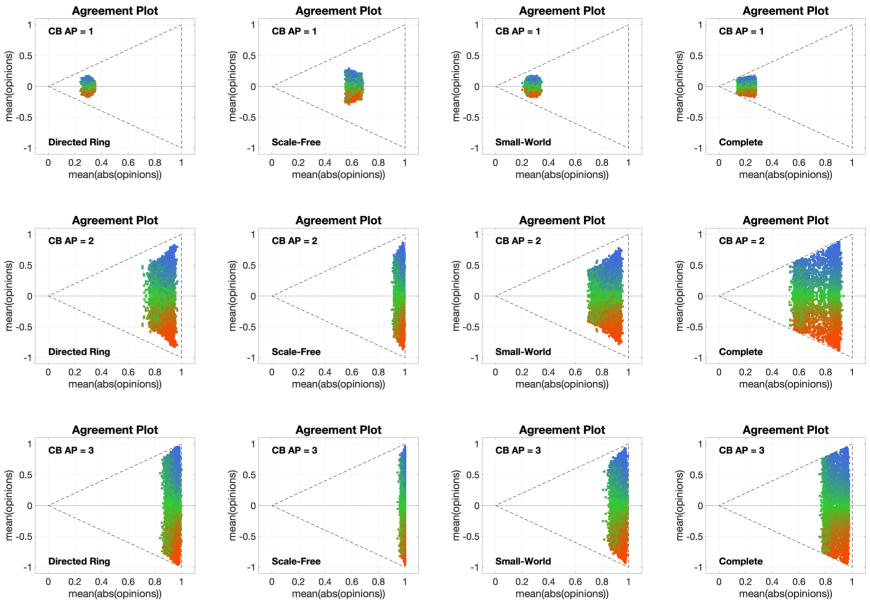


Figure 13: **IOSS** plots for the Classification-based model. Each of the 12 **IOSS** plots includes 5314 points with different choices of initial opinion distributions and constant agent parameters and underlying digraphs. Plots in the same row have the same agent parameters. Plots along the same column have the same underlying digraph, (from left to right, Directed Ring, Scale-Free, Small-World, and Complete). The agent parameter number can be seen in the upper left corner (Fig 9 shows a representation of the corresponding agent parameters). The underlying digraph name can be seen in the bottom left corner, and some metrics are presented in Table 2. All simulations are with 100 agents for 1000 time steps.

Fig 13 shows 12 **IOSS** plots for 3 different agent parameters and 4 different underlying digraphs. These plots are very interesting. They confirm the existence of 'convergence regions' towards which all the opinions move. Clearly these 'converge regions' depend on the agent parameters and underlying digraph, as hinted by Fig 11. These regions are symmetric with respect to reflections along the x-axis, given that the model has no preference for agreement or disagreement. Besides this observation there is no clear relation between the 'convergence region' shape or location and the agent parameters and underlying digraph.

An additional observation is that there are no blue dots with negative opinion mean, and there are no red dots with positive opinion mean. This indicates that the opinion's general agreement does not travel significantly along the y-axis, or in other words that the mean of the opinions does not change much, as previously noted. Finally, we can see that the 'convergence region' is sometimes located away from the $(0,0)$, $(1,-1)$, $(1,1)$ triangle's boundary, which is consistent with the behaviour noticed in Fig 12.

After the Agreement Plot analysis of Figs 10 to 13, we can summarise the behaviour and intrinsic properties of the Classification-based model as follows:

- Regardless of where the initial opinions are located in the Agreement Plot, they will move towards a 'convergence region' which depends on the agent parameters and the underlying digraph. It is possible for the convergence region to be located near the boundaries of the $(0,0)$, $(1,-1)$, $(1,1)$ triangle, but that is not necessary.

- Although the agent parameters have a significant effect in the opinion evolution, the predicted opinion distribution location is also highly dependent on the underlying digraph, much more than in the Friedkin-Johnsen model.

- For the studied underlying digraphs and initial opinions, the collection of possible opinion outcomes for a wide variety of agent parameters forms a connected and possibly convex region in the Agreement Plot that contains the initial point.

- It is rare for predicted opinions to be located near the $y = \pm x$ and $x = 1$, meaning that almost always there are some agents that agree and others that disagree, and that outcomes where all agents have extreme opinions are quite uncommon.

Conclusion

We have proposed a novel analysis technique for agent-based opinion formation models. This technique, called the **Agreement Plot**, represents several opinion evolutions in the same Cartesian plane thus allowing for a thorough and systematic comparison of opinion evolutions for different initial opinions, agent parameters, and underlying digraphs. This in turn, evidences if and how global patterns produced by the model relate to these model parameters and also reveals intrinsic model properties.

We used the **Agreement Plot** to plot parametric curves of the opinion evolution (respectively, predicted opinion distributions) for different agent parameters and initial opinions, resulting in the **APTE** and **IOTE** (resp. **APSS** and **IOSS**) plots. By systematically varying the agent parameters, underlying digraph, and initial opinions of these plots we were able to characterise how the model predictions relate to these model properties, and which patterns are intrinsic to the models and which depend on the chosen parameters.

This analysis was done for the classical Friedkin-Johnsen model, and the recently proposed Classification-based model. For the Friedkin-Johnsen model the Agreement Plot analysis revealed that when the average path length and diameter of the underlying digraph decreases, the opinions change faster, and the final asymptotic value is closer to the line $y = \pm x$; when the mean susceptibility is high, and the average path length and digraph diameter are low, the opinions reach the $y = \pm x$ line and move parallel to it; and the final predicted opinions rarely have a higher mean of the opinion absolute values than the initial opinions, and, if this happens, the opinions are along the $y = \pm x$ line; and in a society with low susceptibility variance, opinions will change more than another society with similar susceptibility mean and higher susceptibility variance.

The presented Agreement Plot analysis for the Classification-based model suggests that opinions will move to a 'convergence region' which depends on the agent parameters and the underlying digraph, regardless of where they start; the final predicted opinions are highly dependent on the underlying digraph, much more than for the Friedkin-Johnsen model; and that outcomes where all agents have extreme opinions are rare.

The proposed technique can be applied to any agent-based opinion formation model in which the agent's opinions belong to a bounded interval in the real line. Overall, the presented technique provides a different perspective to study and analyse opinion formation models. In a graphical and intuitive way it shows how the opinion evolution depends on the model parameters and also what the model capabilities are. This is in part thanks to the representation of complete opinion distributions by a single point in the Cartesian plane. Although some information is lost, the most relevant information is kept and highlighted. The proposed methodology can be modified to other analysis interest. For instance, a third dimension could be added, or the colour coding could be changed, to represent an additional relevant metric.

Supporting information

SI Supplementary Information Additional document providing an explanation on how to reproduce the results presented in this paper; a description of the Classification-based model; and explanation on how the network metrics were computed [PDF]

Acknowledgments

The ternary plots were made using modified scripts from Carl Sandrock [35].

References

[1] J. French Jr., "A formal theory of social power," *Psychological Review*, 1956.

[2] F. Harary, "A criterion for unanimity in french's theory of social power," *Studies in social power*, 1959.

[3] F. Harary, a. Harary, R. Norman, D. Cartwright, and K. Esau, *Structural Models: An Introduction to the Theory of Directed Graphs*. No. vol. 82, Wiley, 1965.

[4] M. Degroot, "Reaching a consensus," *Journal of the American Statistical Association*, 1974.

[5] K. Kacperski and J. Holyst, "Opinion formation model with strong leader and external impact: A mean field approach," *Physica A: Statistical Mechanics and its Applications*, vol. 269, no. 2, pp. 511–526, 1999.

[6] K. Kacperski and J. Holyst, "Phase transitions as a persistent feature of groups with leaders in models of opinion formation," *Physica A: Statistical Mechanics and its Applications*, vol. 287, no. 3-4, pp. 631–643, 2000.

[7] P. Sobkowicz and A. Sobkowicz, "Dynamics of hate based internet user networks," *European Physical Journal B*, vol. 73, no. 4, pp. 633–643, 2010.

[8] A. Chmiel, P. Sobkowicz, J. Sienkiewicz, G. Paltoglou, K. Buckley, M. Thelwall, and J. Hoyst, "Negative emotions boost user activity at bbc forum," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 16, pp. 2936–2944, 2011.

[9] R. Hegselmann and U. Krause, "Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model," *Networks and Heterogeneous Media*, vol. 10, no. 3, pp. 477–509, 2015.

[10] N. Masuda, "Opinion control in complex networks," *New Journal of Physics*, vol. 17, pp. 1–11, 2015.

[11] P. Sobkowicz, "Opinion dynamics model based on cognitive biases of complex agents," *Journal of artificial societies and social simulation*, vol. 21, no. 4, 2018.

[12] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5791–5796, 2013.

[13] S. Banisch and H. Shamon, "Biased processing and opinion polarisation: experimental refinement of argument communication theory in the context of the energy debate," *Available at SSRN 3895117*, 2021.

[14] J. Lorenz, M. Neumann, and T. Schröder, "Individual attitude change and societal dynamics: Computational experiments with psychological theories.," *Psychological Review*, vol. 128, no. 4, p. 623, 2021.

[15] X. Yin, H. Wang, P. Yin, and H. Zhu, "Agent-based opinion formation modeling in social network: A perspective of social psychology," *Physica A: Statistical Mechanics and its Applications*, vol. 532, 2019.

[16] M. Krawczyk, K. Malarz, R. Korff, and K. Kułakowski, "Communication and trust in the bounded confidence model," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6421 LNAI, no. PART 1, pp. 90–99, 2010.

[17] N. Friedkin, "A formal theory of social power," *Journal of Mathematical Sociology*, 1986.

[18] N. Friedkin and E. Johnsen, "Social influence networks and opinion change," *Advances in Group Processes*, vol. 16, 01 1999.

[19] E. Chattoe-Brown, "Using agent based modelling to integrate data on attitude change," *Sociological Research Online*, vol. 19, no. 1, pp. 159–174, 2014.

[20] P. Duggins, "A psychologically-motivated model of opinion change with applications to american politics," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 1, p. 13, 2017.

[21] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of artificial societies and social simulation*, vol. 5, no. 3, 2002.

[22] S. Schweighofer, F. Schweitzer, and D. Garcia, "A weighted balance model of opinion hyperpolarization," *Journal of Artificial Societies and Social Simulation*, vol. 23, no. 3, p. 5, 2020.

[23] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, "Modeling echo chambers and polarization dynamics in social networks," *Physical Review Letters*, vol. 124, no. 4, p. 048301, 2020.

[24] G. Fu and W. Zhang, "Opinion formation and bi-polarization with biased assimilation and homophily," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 700–712, 2016.

[25] J. Su, B. Liu, Q. Li, and H. Ma, "Coevolution of opinions and directed adaptive networks in a social group," *Journal of artificial societies and social simulation*, vol. 17, no. 2, 2014.

[26] P. Sobkowicz, "Studies of opinion stability for small dynamic networks with opportunistic agents," *International Journal of Modern Physics C*, vol. 20, no. 10, pp. 1645–1662, 2009.

[27] M. Meadows and D. Cliff, "Reexamining the relative agreement model of opinion dynamics," *Journal of Artificial Societies and Social Simulation*, vol. 15, no. 4, p. 4, 2012.

[28] T. Kurahashi-Nakamura, M. Mäs, and J. Lorenz, "Robust clustering in generalized bounded confidence models," *Journal of Artificial Societies and Social Simulation*, vol. 19, no. 4, p. 7, 2016.

[29] H. Liang, Y. Dong, and C. Li, "Dynamics of uncertain opinion formation: An agent-based simulation," *Journal of Artificial Societies and Social Simulation*, vol. 19, no. 4, p. 1, 2016.

[30] C. A. Devia and G. Giordano, "A framework to analyze opinion formation models," *Scientific Reports*, vol. 12, no. 1, p. 13441, 2022.

[31] C. A. Devia and G. Giordano, "Probabilistic analysis of agent-based opinion formation models," 2023.

[32] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz, "Models of social influence: Towards the next frontiers," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 4, 2017.

[33] S. Galam, "Opinion dynamics and unifying principles: A global unifying frame," *Entropy*, vol. 24, no. 9, p. 1201, 2022.

[34] C. A. Devia and G. Giordano, "Classification-based opinion formation model embedding agents' psychological traits," 2022.

[35] C. Sandrock, "alchemyst/ternplot." Available at `https://www.mathworks.com/matlabcentral/fileexchange/2299-alchemyst-ternplot?s_tid=srchtitle` (2022/06/13), 2012.