



HAL
open science

Modeling both pairwise interactions and group effects in polarization on interaction networks

Duncan Cassells, Lionel Tabourier, Pedro Ramaciotti

► To cite this version:

Duncan Cassells, Lionel Tabourier, Pedro Ramaciotti. Modeling both pairwise interactions and group effects in polarization on interaction networks. International Conference on Complex Networks, Apr 2024, Exeter, United Kingdom. hal-04363001

HAL Id: hal-04363001

<https://hal.science/hal-04363001>

Submitted on 23 Dec 2023

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.

Modeling both pairwise interactions and group effects in polarization on interaction networks

Duncan Cassells^{1,2,4}, Lionel Tabourier¹, and Pedro Ramaciotti^{3,2,4}

¹ LIP6, Sorbonne Université

² LPI, Université Paris Cité

³ Complex Systems Institute of Paris Île-de-France CNRS

⁴ médialab, Sciences Po

duncan.cassells@lip6.fr

Abstract. The study of polarization has gained increasing attraction in the past decades. Since observing both opinions and interactions is challenging, epistemic programs such as agent-based models have been proposed as a means to assessing the systemic consequences of social psychology mechanisms. Most results in agent-based models for opinion dynamics have focused on individual opinion constructs and pairwise interactions, with a few works treating group effects as constraints. Meanwhile, a tradition in social sciences has been putting emphasis on how group configuration affects individual behavior. In this work, we introduce a new model for accounting for both pairwise interactions in which actors observe and update opinions, and individual perception of the evolving configuration of groups that make up the population in which they are embedded. Through experiments, we show that different treatment given to pairwise interactions, depending on whether they occur between in-groups or out-groups, has quantifiable impacts in the resulting polarization of a population. In particular, the tolerance toward out-group opinions is shown to have a strong impact on the resulting polarization. Our model produces and accounts for polarized states resulting from group consolidation and fragmentation.

Keywords: opinion dynamics · polarization · group effects.

1 Introduction

Polarization has become a central topic of interest among social scientists [9, 12]. The public perceives society as more and more polarized [17] and political elites have been shown to be also increasingly polarized [21]. The study of polarization is challenging for several reasons. Firstly, it hinges on the notion of individual opinion, which is difficult to observe and measure jointly with interactions. Further, the conceptualization and operationalization of polarization offers several challenges, as they involve at least notions of attitudes or opinions towards issues [10], ideologies or structures of opinions and beliefs [6], and social group dynamics [5, 23]. These challenges have seen proposals of different epistemic programs,

including agent-based models for the study of opinion dynamics (see [16, 18] for extensive surveys on this stream of research).

In most agent-based models for opinion dynamics, opinion change is modeled as a result of pairwise interactions, with the opinion of an actor depending on its own opinion, and on the opinion of those with whom it interacts. Only a small minority of works addresses the question of group effects: how opinion change depends on the configuration of groups in the population of actors. At the same time, decades of research in sociology and social psychology have shown that we evaluate others, both individually and according to their group membership, based on a diverse set of cues [13]. The literature on group opinion dynamics has mainly focused on groups as constraints [7, 3], and less on the question of group perception and the elements on which this perception may be predicated. In this article, we present a model for accounting for the perception of groups on the part of the actors, and the results this entails for opinion polarization.

Because we aim at assessing the role of group perception in opinion models, our starting point are models in which opinion update operates by comparison to the opinion of those with which an actor interacts. This means that we intentionally leave outside this work other mechanisms that have been proposed, such as nonlinear opinion updating, asynchronous interactions, or utility functions (please refer to [18] for a proposed taxonomy of these elements). We consider for our model the attraction-repulsion model (ARM) [1]. While relying on linear opinion updating depending on pairwise interactions, the ARM also allows for the possibility of interactions resulting in opinion distancing. This mechanism has been shown to be relevant in specific settings (*e.g.*, in online social networks [2]) and has been hypothesized to be caused by motivated reasoning increasing the perceived differences in stances between already distant opinions [15, 22].

In this article we build on the ARM to propose a novel mechanism for accounting for the perception of groups on the part of actors. The ARM model hinges on three parameters: responsiveness R (the degree of opinion change as a proportion of the opinion distance between two actors), exposure E (modulating the probability that two actors will interact), and tolerance T (a threshold on the value of opinion distance between two actors, beyond which, the opinion change resulting from interaction will separate them in the opinion scale, as opposed to attracting them). Based on the aforementioned research [15, 22] highlighting how attitudes towards members of the same group (*i.e.*, in-groups) differ from attitudes towards actors outside the same group (*i.e.*, out-groups), our model considers two different values of tolerance thresholds: one for in-groups T_{in} , and one for out-groups T_{out} . The perception of groups in our model is shared by all actors, and is computed on the basis of the degree to which different actors display opinions that are spatially clustered in the opinion space. We use our model with different parameters, running simulations and showing opinion convergence towards states displaying different degrees of polarization. We evaluate polarization as a property of the spatial distribution of opinion of the population in our model, using the Duclos-Esteban-Ray measure [8], which accounts for both spread and local clustering. Our results show that different tolerance

values for in- and out-groups play an important role in describing different ways by which populations might polarize. Within this model, we observe that the attitude of agents towards out-groups impacts the model strongly, while the attitude towards in-groups is nearly indifferent, except for a low in-group tolerance. The results are also consistent with a mechanistic plausible explanation for the polarization of groups by means of fracturing from within of opinion groups, and conversely, the consolidation of two groups to reduce polarization.

2 Model Description

Our model consists of a population of $N = 100$ actors, each with an opinion represented by a numeric value on a closed interval between the extremes 0 and 1. Actors may then change their opinions in successive iterations until the simulation reaches a stable state.

At each iteration of the model, each actor $i \in \{1, \dots, N\}$ interacts with a subset of the rest of the population (decided by an interaction rule) and so observes the opinions of these other actors. Based on these observations the actor i updates their opinion. Both the interaction rule and opinion updating process are described below in Sections 2.1 and 2.2, respectively.

The stability criterion that we impose is that our measurement of polarization changes over the previous twenty iterations by less than 2%. This occurs within at most 200 iterations across all cases in this study. We run all of our simulations for 300 iterations for completeness.

We measure polarization using the Duclos-Esteban-Ray measure [8], detailed in Section 2.3, as well as standard deviation to provide comparison with previous works and reference. This gives us a single value metric to assess polarization across our entire population.

2.1 Interaction Rule

All actors may potentially connect with all other actors, *i.e.*, we have a complete social graph. Interactions may occur on these connections with probability $P(i \rightarrow j) = (1/2)^{d_{ij}/E}$ (the probability that i observes the opinion of j), where d_{ij} is the absolute distance between the opinions o_i and o_j on $[0, 1]$ of actors (*i.e.*, nodes) i and j , respectively, and E is a model parameter between 0 and 1 which represents *exposure* – governing whether interactions with distant actors are more, or less, likely. This interaction rule reflects the principle of homophily [20] (and more specifically *value homophily* [14]), given that there is an increased tendency to interact with others who are more ideologically similar.

Value E determines, at each iteration, an interaction network within the complete social graph. Higher values of E increase the probability of interaction for all non-zero distances on the $[0, 1]$ interval. Fig. 1 illustrates the probability of pairwise interactions as a function of distance d_{ij} and E .

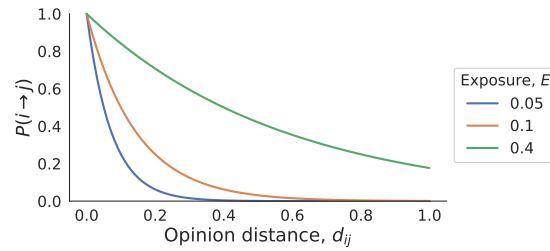


Fig. 1: The effect of parameter E (exposure) in determining the interaction network on top of the complete social graph, by setting the probability of pairwise connections as a function of their opinion distance.

2.2 Opinion Updating Process

The model that we use for determining opinion change is an Attraction-Repulsion Model (ARM), also termed as assimilative-repulsive social influence [1, 11] which has recently gained more attention due to results showing that actors can be repulsed by opposing views while being attracted to similar ones [2].

There are two options for the opinion updating rule in the model, dependent on a *tolerance* threshold value T , between 0 and 1: the first option is that when an actor interacts with another actor of similar opinion ($d_{ij} \leq T$) then the actor’s opinion changes by moving towards the other’s – the actor is *attracted*; the second is the opposite in that if actors are sufficiently dissimilar ($d_{ij} > T$) then opinion difference increases – the actor is *repulsed* by the other.

There is also a second parameter for opinion updating, which governs the magnitude of the attraction or repulsion and is set by value R representing *responsiveness* such that the change in opinion is $\pm R \cdot d_{ij}$ (with the sign depending on the comparison of distance d_{ij} tolerance threshold T). Responsiveness can take values between 0 and 1 depending on the fraction of opinion difference which is to be covered.

We introduce the notion of groups, meaning that each actor may belong to one of many groups on which the population is assorted. In this work we extend the standard ARM described so far to propose a tolerance threshold that is dependent on whether the interaction occurring is in-group or out-group, that is whether the two actors have the same group identity (*i.e.*, belong to the same group), or not.

Group identity is determined by the HDBScan clustering algorithm [19] with a minimum group size of $N/5 = 20$ to avoid unnecessary noise within our population, and computed at each iteration (group configuration can change in time). If actors do not clearly belong to a group (as decided by the algorithm) then they are left without a group identity and so are treated as out-group by defined groups [24] (*i.e.*, actors without group will always perceive, and be perceived by others, as out-groups).

The modeling choice to propose group-dependent tolerance is motivated by two observations from the state of the art. First, the authors of the ARM model note that polarization outcomes are largely determined by tolerance T [1]. Second, as discussed in the introduction, opinion dynamics are as much a result of psychological constructs such as opinion as they are determined by social dynamics such as group configuration. We thus split T into T_{in} and T_{out} in our model. If an interaction happens between two actors in the same group, the value of tolerance will be taken to be T_{in} , else, it will be taken to be T_{out} . Table 1 summarizes the main parameters of our model, comparing them to the ARM model.

Table 1: Comparison with the Attraction-Repulsion model [1].

Model Element	ARM	This Paper	Considerations
<i>Iteration Step</i>	One pair of actors is randomly selected, then subject to the interaction and update rules.	All actor pairs are subject to the interaction and update rules.	Fewer iterations to reach stability but equivalent final states reached.
<i>Parameters</i>	E, R, T	E, R, T_{in} , and T_{out}	Tolerance parameter is group-dependent.
<i>Actors</i>	$N \sim \mathcal{N}(0.5, 0.2^2)$	$N/2 \sim \mathcal{N}(\frac{1}{3}, 0.05^2)$, $N/2 \sim \mathcal{N}(\frac{2}{3}, 0.05^2)$	Both are normally distributed, but we consider Gaussian Mixture Models at $t = 0$ to account for group configurations.

2.3 Polarization Metrics

There are currently a variety of polarization measures used in the literature [4]. We use the Duclos-Esteban-Ray (DER) measure [8]. We prefer to use this rather than alternatives because it provides a single value measure derived from an axiomatic theory specifically for polarization, accounting for both spread (as in other commonly used polarization metrics, such as standard deviation) and local concentration (accounting for multi-modality in the opinion distribution of the population).

The form of the DER polarization measure is

$$P_\alpha(f) \equiv \int \int f(x)^{1+\alpha} f(y) |y - x| dy dx, \quad (1)$$

where $\alpha \in [1/4, 1]$ and we set $\alpha = \frac{1}{2}$ for this study (see [8, Section 3.2] for a sensibility analysis). We use a sample based estimator for the measure which is given by [8, Section 4]:

$$P_\alpha(\hat{F}) = \frac{1}{n} \sum_{i=1}^n \hat{f}(x_i)^\alpha \hat{a}(x_i), \quad (2)$$

with $\hat{a}(x_i)$ given as

$$\hat{a}(x_i) = \hat{\mu} + x_i \left(\frac{(2i-1)}{n} - 1 \right) - \frac{1}{n} \left(2 \sum_{j=1}^{i-1} x_j + x_i \right), \quad (3)$$

where $\hat{\mu}$ is the sample mean, and the estimated underlying density function $\hat{f}(\cdot)$ is found by kernel density estimation.

The Pearson correlation coefficient between DER and standard deviation measured on the final stable states of over 400 simulations is 0.99. One might well question why we do not use the better known standard deviation then, but we opt for DER as we consider it a more robust measure for the previously mentioned reasons. Furthermore, given that DER can deal well with multi-modality as an aspect of polarization - unlike standard deviation - this robustness should bear well under further extensions of the model.

3 Results

Simulation process. Our initial conditions for each simulation on the population of 100 actors is that they are equally divided into two opinion groups at a starting iteration $t = 0$. This gives us a bimodal distribution along the $[0,1]$ interval of our opinion space - this is chosen to align with notions of ideological polarization.

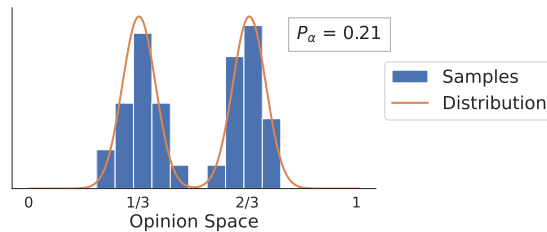


Fig. 2: An example initial distribution of the population’s opinions at $t = 0$. We also display the DER measurement, P_α , of the distribution.

To create the bimodal distribution we draw from a Gaussian Mixture Model comprised of two densities with differing means. More precisely, half of the population’s opinions are randomly assigned $\sim \mathcal{N}(\frac{1}{3}, 0.05^2)$, with the other half $\sim \mathcal{N}(\frac{2}{3}, 0.05^2)$. An instance of a sample drawn from the distribution created by the combination of these two Normal distributions and the samples generated

for initial conditions can be seen in Fig. 2. The initial population is always drawn from the same underlying distribution but the samples are not always identical given the random draws.

From the initial iteration at $t = 0$, we update the population through iterations for $t \in \{1, \dots, 300\}$. At the start of each iteration step t_i the groups are identified using the clustering algorithm HDBScan. With group identity assigned we call on the interaction rule, once these information exchanges have been determined, we then call on the opinion updating process to act on the observed opinions from the interaction phase. We run these iterations until the population reaches a steady polarization state (*i.e.*, when change in polarization P_α is less than 2% over the previous twenty iterations) as described above. We term this collection of iterations as one simulation, and note that for each simulation the parameters E , R , T_{in} , and T_{out} are constant.

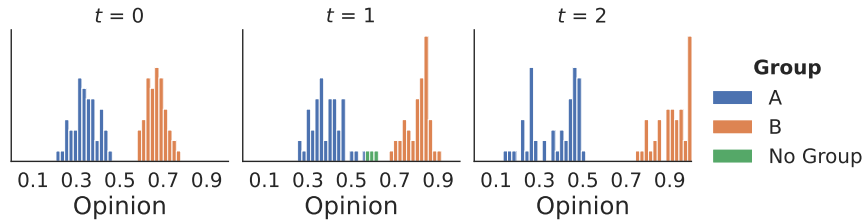


Fig. 3: The evolution of opinion and group identity during the first three iterations for a simulation with $E = 0.1$, $R = 0.1$, $T_{\text{in}} = 0.6$, and $T_{\text{out}} = 0.55$.

A snapshot of how a couple of these iterations develop is shown in Fig. 3. We see how the distribution changes iteration to iteration as determined by the interaction and update rules, and how groups are then adjusted depending on the new opinions of the population.

Assessing model behaviour. To appraise the impact that group-dependent tolerances have on the evolution of polarization, we run simulations across a range of values for T_{in} and T_{out} , while keeping E and R constant. Additionally, we present the simulations under different exposure and responsiveness conditions, but a deeper exploration of the role of these parameters can be found in Axelrod et al.’s original paper [1] given that our focus is on the group dependent tolerances. Under the case of no differentiation between in-group and out-group ($T_{\text{in}} = T_{\text{out}}$), our model arrives at equivalent results to their conclusions.

For the final state polarization value of a simulation we take the mean of the final ten iterations, since each simulation is considered stable after meeting the stability criteria, but polarization values are not absolutely constant. To illustrate the convergence to these stable values Fig. 4 shows the trajectory of the mean DER value over the fifty simulations - since the interaction involves a

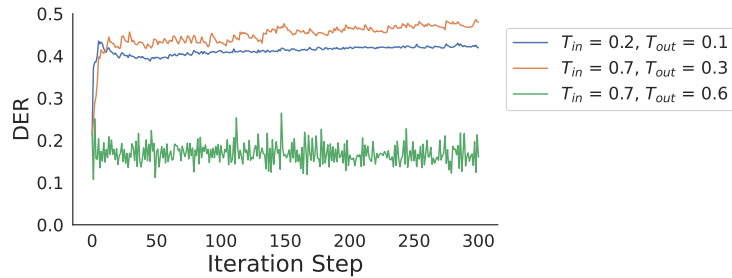


Fig. 4: Convergence to stable DER values for various $T_{\text{in}}/T_{\text{out}}$, and $E=0.1$, $R=0.1$.

non-deterministic procedure, we run fifty simulations for each parameter combination and present the mean result of these as an average case polarization outcome. Some cases converge to stability relatively quickly while the mid-range polarization values can take up to nearly two hundred iterations to achieve stability.

To evaluate the impact of parameters on each simulation (particularly of our new parameters T_{in} and T_{out}) we show the polarization of simulations across a range of parameter values in Fig. 5. The central result of this work is that by introducing group sensitivity into the attraction-repulsion model we observe polarization behaviours that we would otherwise miss. This is clear in observing that the diagonal in Fig. 5 is the case where $T_{\text{in}} = T_{\text{out}}$, and as such it is the special case where group identity has no role in tolerance. When we move away from the diagonal of the grid we observe variation in the measure of polarization which can only be caused by group dependent tolerances, thus indicating the influence of the group effects. We also note that the upper-left triangle of Fig. 5 - where $T_{\text{in}} \geq T_{\text{out}}$ - is most meaningful in real world context (since it means that actors are more tolerant to others inside their group than outside their group), but the exploration of the full space relevant to full model understanding. We further explain the intricacies of the $T_{\text{in}}/T_{\text{out}}$ parameter space below.

Considering the simulations where $E = 0.1$ and $R = 0.1$ (Fig. 5a) we observe approximately three zones: the upper left (*i.e.*, high T_{in} and low T_{out}) which is highly polarized (DER > 0.3), the upper right (*i.e.*, high T_{in} and high T_{out}) which has mid-to-low level polarization (DER < 0.2), and the region below $T_{\text{in}} \simeq 0.2$. The variable polarization observed in this last region of the parameter space can be explained by the increased tendency of groups to fracture from within at low T_{in} , as a result of pairwise distancing between actors inside a group. The vertical boundary between upper left and upper right regions is where consolidation between out-groups starts to occur. That is, when T_{out} is sufficiently large, interactions with out-group actors start to become increasingly more attractive. We next provide further discussion of these zones.

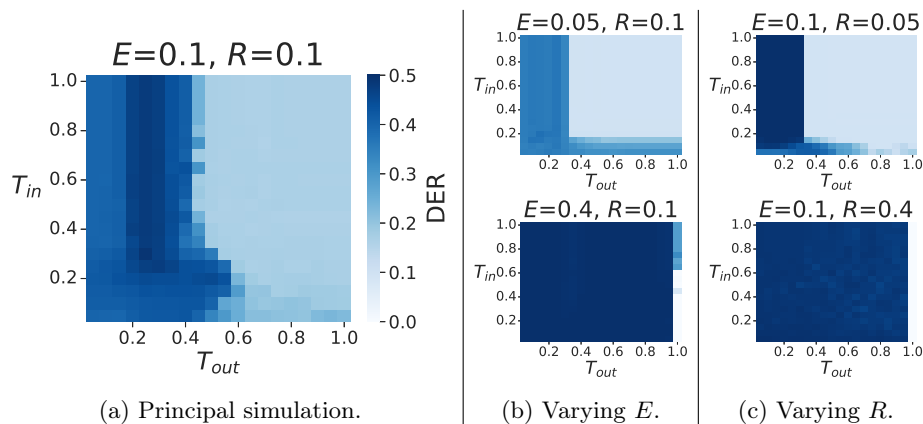


Fig. 5: Each grid shows the stable mean DER across 50 simulations with varying T_{in} and T_{out} , and constant E and R .

Consolidation with out-group due to T_{out} . For fixed E and R , we observe in Fig. 5 that above $T_{in} \simeq 0.2$ the threshold between high and low polarization depends on T_{out} only. This asymmetry between the two parameters suggests that, except for the case of low tolerance within group, the qualitative behaviour is essentially governed by the tolerance to outsiders.

The boundary between which low polarization ($DER < 0.2$) and high polarization ($DER > 0.3$) occurs is that point at which the two groups are able to tolerate at least some of the other group’s actors’ opinions (we also note that the exact position of the boundary is influenced by E and R too).

At low $T_{out} (\leq 0.2)$ the steady state polarization is lower than the region $0.2 < T_{out} \leq 0.4$. This is due to T_{out} being so low that the formation of two groups from the distribution, one at either extreme of our scale for max polarization, does not occur. Instead the result is a group at one extreme, a group at the other extreme, *and* a group that floats between them (repulsed by the two extreme groups). We see this in Fig. 6b (displaying the average number of groups at iteration stages) where the low T_{out} region is darker, indicating on average 3 or more groups, and this pattern continues in further iterations.

Fracturing of in-groups due to T_{in} . This second polarizing phenomenon occurs when in-group tolerance is sufficiently low - below the threshold of $T_{in} \simeq 0.2$. In this case actors can no longer tolerate some other opinions within their group and so this drives fracturing of the group from within as repulsion happens.

We provide a view of the average number of groups in the first ten iterations in Fig. 6a. The number of groups identified during the first iterations show us that low T_{in} fractures in-groups and this happens quickly since interactions close in opinion space (within group) occur more frequently. As such we end with > 2 groups in our opinion space as the original groups break down. Note that for

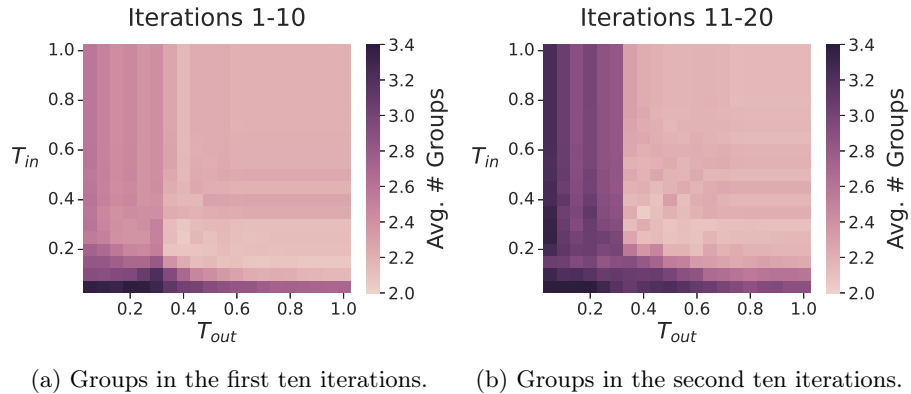


Fig. 6: Average number of groups identified within the distribution are shown for $E = 0.1$, $R = 0.1$ and across T_{in} and T_{out} , for the first 1-10, or 11-20, iterations.

these measures, we count the existence of actors left unidentified by the algorithm as a group (occurring due to the minimum group size condition) so that we can observe and quantify small groups appearing between the two larger groups.

T_{in} being so low that it causes fracturing may seem like a non-standard situation as in this case how would groups form in the first place, so it is clearly not relevant in the formation of groups. However, if existing tolerance within group were to reduce due to some external factor then this becomes conceivable.

Impact of other model parameters on the tolerance space. While group dependent tolerances are the focus of this work, exposure and responsiveness parameters clearly still have impact on the polarization outcomes of the system, as seen in comparing across the different parameter grids in Fig. 5. They moderate sharpness and exact placement of boundaries in the state space.

If exposure E is low then high polarization does not occur, since actors are rarely exposed to a significantly different opinion and so we do not observe strong repulsion between groups. On the other hand, high exposure makes interaction with lots of different opinions more likely and so encourages actors to move a lot around the opinion space, which brings them to the boundaries. Boundary positions are only broken if tolerance is so high that any opinion is tolerated and thus every interaction is attractive, which is a parameter boundary case of the model.

Increasing R varies the rate of exploration of the space, by allowing wilder swings of opinion. This also pushes actors to the edge of the state space and so results in high polarization. Low R creates sharp boundaries between regions since actors move more slowly and thus there is reduced noise in the outcomes. The reader is referred to the original ARM paper [1] for an in-depth analysis on the impact of parameters E and R .

4 Conclusions

In this work, we have presented the value of including group identity within the field of opinion dynamics work, particularly in extending previous work on attraction-repulsion type models. Given the significant interest in the role of group identity in polarization from a political science and sociological perspective, we were interested in showing that modelling group effects leads to new, potentially unexpected, qualitative behaviours in regards to polarization.

We identified out-group tolerance to be a key part of understanding the development of polarization as non-consolidation of groups; although we keep in mind that sufficiently reasonable exposure and responsiveness conditions also play a role in where the exact boundary between high and low polarization exists. Secondly, it was shown that if tolerance of differing opinions within a group is not high enough then the group will fracture and thus contribute to higher levels of polarization.

In future work, we suggest to evaluate the robustness of the model results to other design choices. For instance, do the main observations stand if we differentiate the perceptions of out-groups or choose an individual-based perception of groups. Then, we will propose using real-world interactional data on populations with known group identification in social media to test the hypotheses brought forward in this work, and to guide the development of inference methods addressing the need to identify the social or opinion traits on which individuals judge group membership.

Acknowledgements This work has been funded by the “European Polarisation Observatory” (EPO) of the CIVICA Research (co-)funded by EU’s Horizon 2020 programme under grant agreement No 101017201. P.R. acknowledges support by the Data Intelligence Institute of Paris (diiP) through the French National Agency for Research (ANR) grant ANR-18-IDEX-0001 “IdEx Université de Paris” and SoMe4Dem (Grant No. 101094752) Horizon Europe project.

References

1. Axelrod, R., Daymude, J.J., Forrest, S.: Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences* **118**(50), e2102139118 (2021)
2. Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A.: Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* **115**(37), 9216–9221 (2018)
3. Berger, R.L.: A necessary and sufficient condition for reaching a consensus using deGroot’s method. *Journal of the American Statistical Association* **76**(374), 415–418 (1981)
4. Bramson, A., Grim, P., Singer, D.J., Fisher, S., Berger, W., Sack, G., Flocken, C.: Disambiguation of social polarization concepts and measures. *The Journal of Mathematical Sociology* **40**(2), 80–111 (2016)

5. Brown, R.: Social identity theory: Past achievements, current problems and future challenges. *European journal of social psychology* **30**(6), 745–778 (2000)
6. Converse, P.E.: *The nature of belief systems in mass publics. Ideology and Discontent*, (1964)
7. DeGroot, M.H.: Reaching a consensus. *Journal of the American Statistical association* **69**(345), 118–121 (1974)
8. Duclos, J.Y., Esteban, J., Ray, D.: Polarization: concepts, measurement, estimation. In: *The Social Economics of Poverty*, pp. 54–102. Routledge (2006)
9. Fiorina, M.P., Abrams, S.J.: Political polarization in the american public. *Annu. Rev. Polit. Sci.* **11**, 563–588 (2008)
10. Fishbein, M., Raven, B.H.: The ab scales: An operational definition of belief and attitude. *Human relations* **15**(1), 35–44 (1962)
11. Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., Lorenz, J.: Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation* **20**(4) (2017)
12. Iyengar, S., Westwood, S.J.: Fear and loathing across party lines: New evidence on group polarization. *American journal of political science* **59**(3), 690–707 (2015)
13. Lamont, M., Molnár, V.: The study of boundaries in the social sciences. *Annual review of sociology* pp. 167–195 (2002)
14. Lazarsfeld, P.F., Merton, R.K.: Friendship as a social process: A substantive and methodological analysis. In: Berger M., A.T., H., C. (eds.) *Freedom and Control in Modern Society*, vol. 18, pp. 18–66. New York, Van Nostrand (1954)
15. Lord, C.G., Ross, L., Lepper, M.R.: Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology* **37**(11), 2098 (1979)
16. Lorenz, J.: Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C* **18**(12), 1819–1838 (2007)
17. Mason, L.: *Uncivil agreement: How politics became our identity*. University of Chicago Press (2018)
18. Mastroeni, L., Vellucci, P., Naldi, M.: Agent-based models for opinion formation: A bibliographic survey. *IEEE Access* **7**, 58836–58848 (2019)
19. McInnes, L., Healy, J., Astels, S.: hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2**(11), 205 (2017)
20. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. *Annual review of sociology* **27**(1), 415–444 (2001)
21. Poole, K.T., Rosenthal, H.: *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand (2000)
22. Taber, C.S., Lodge, M.: Motivated skepticism in the evaluation of political beliefs. *American journal of political science* **50**(3), 755–769 (2006)
23. Tajfel, H., Turner, J.C.: The social identity theory of intergroup behavior. In: *Political psychology*, pp. 276–293. Psychology Press (2004)
24. Yang, V.C., van der Does, T., Olsson, H.: Falling through the cracks: Modeling the formation of social category boundaries. *Plos one* **16**(3), e0247562 (2021)