Applied Network Science

RESEARCH Open Access

The power of social networks and social media's filter bubble in shaping polarisation: an agent-based model

Cristina Chueca Del Cerro[1*]

*Correspondence: cristina.chueca-del-cerro@durham.ac.uk

[1] Department of Sociology, Durham University, Durham, UK

Abstract

Tthe role social media platforms play on the emergence of polarisation is an ongoing debate in the political communication literature. Social media's filter bubbles and online echo chambers shape people's opinions by curating the information they have available. However, the extent to which this is the case remains unclear. Social simulation scholars have provided valuable insights into the subject through opinion dynamics models and agent-based modelling approaches. This article proposes a social simulation approach to the topic of opinion dynamics from a political communication perspective to understand how social network configurations and the media environment contribute to the emergence of national identity polarisation. We built an agent-based simulation model of national identity dynamics with a multilayer multiplex network of interacting agents in a hybrid media environment of both, traditional media and social media platforms. We use the Catalan secessionist movement to ground, contextualise and empirically inform parts of our model. We found that the initial social network setup conditions had a large impact on the emergence of polarisation amongst agents. In particular, homophily-based social networks composed of a majority of like-minded individuals produced greater polarisation compared to random networks. This was aggravated in the presence of social media filtering algorithms, selectively exposing agents to supportive information. These results emphasise the importance of both the selective exposure by social media filtering algorithms and one's social networks (echo chambers) for polarisation to emerge. This interaction reinforces the influence of social media platforms and social networks have on the emergence of polarisation.

Keywords: Social networks, Social media, Agent-based model, Polarisation

Introduction

Social media platforms have changed the way we communicate with each other as well as the way we engage in politics. The effects of the social media environment have on the polarisation of political attitudes are varied (Lelkes 2016; Dubois and Blank 2018; Iyengar et al. 2012; Tucker et al. 2018). Social media platforms act as an amplifier for polarisation due to the unrestricted discussions and commentary. Added to this is the high connectivity between individuals and the anonymity granted in certain platforms.

Studies have found that discussions of political issues online tend to polarize the users involved (see Smith et al. 2014; Tucker et al. 2018; Li and Xiao 2017). However, numerous studies have pointed out that such claims might be overstated and that online polarisation is either smaller than originally posited or non-existent at all (Osmundsen et al. 2021; Tucker et al. 2018; Guess 2021; Bakshy and Messing 2015; Dubois and Blank 2018; Goel et al. 2010). The lack of consensus in the literature has shifted the focus away from the mechanisms responsible for promoting attitude and opinion change.

One way in which it has been argued social media platforms shape opinions and can promote polarisation is through filter bubbles (Pariser 2011). Social media platforms select the information users are exposed to via their sorting algorithms. These filter through the large amounts of information available and present the information expected to be of greatest interest to the user (Iyengar and Hahn 2009). Filter bubbles operate differently for each social media platform (Bozdag 2013; Bandy and Diakopoulos 2021) but all follow the principles of selective exposure and personalisation. As people are exposed to multiple sources, they select a few of them and interact with them. The filter bubbles initially limit the user's exposure to cross-cutting content (Maes and Bischofberger 2015), content that might not agree with the user's viewpoint. This technology-induced selective exposure of filter bubbles is problematic as it can reinforce existing views (Trilling and Schoenbach 2013; Ross-Arguedas et al. 2022). Therefore, contributing to online attitude polarisation (Flaxman et al. 2016), as shown by empirical research (Levendusky 2013). Yet, some empirical studies using survey data or passive tracking data disagree regarding the extent to which filter bubbles limit online user's exposure to diverse sources (see Dubois and Blank 2018; Flaxman et al. 2016; Fletcher et al. 2021a; Beam et al. 2018).

Echo chambers are the result of that self-selective exposure and selective avoidance of certain information and people. These are user-made homophilous clusters where most users share the same views about an issue and where individuals primarily share information that supports their views. Individuals tend to seek out information that supports their beliefs while being critical of information that is contrary to their beliefs (see Lodge and Taber 2000; Cioroianu et al. 2018; Tucker et al. 2018). Confirmation bias is the concept used to describe this selective exposure and selective avoidance (Festinger 1957). Numerous studies have found that individuals do indeed selectively expose themselves to information, with varying degrees (see Kim and Lu 2020; Garrett and Stroud 2014; Del Vicario et al. 2017). This process of selective exposure and selective avoidance is relevant as it helps explain the emergence of polarisation.

Individuals select political sources that agree with them while minimizing the effort of choosing between sources (Vaccari et al. 2016). Not only individuals will selectively expose themselves to information that supports their views but will also be more likely to believe such information and regard contrary information more negatively. This is also known as disconfirmation bias (Taber and Lodge 2006, see Pennycook and Rand 2019 for review on this topic). Experimental research has demonstrated that indeed, individuals are more likely to select opinion-consistent sources, especially if they hold strong ideological positions or are more politically interested (Iyengar and Hahn 2009; Kim and Lu 2020; Garrett et al. 2013). Furthermore, only getting exposed to information that supports one's views has been found to reinforce them (Axelrod 1997; Levendusky 2013; Stroud 2008).

Yet, certain authors have pointed out that the extent to which echo chamber exist or are effective may have been overstated (Osmundsen et al. 2021; Tucker et al. 2018; Guess 2021; Boutyline and Willer 2016). They argue that individuals have more autonomy over which sources to get exposed to and that citizens tend to seek out alternative opinions instead of choosing to remain solely exposed to supportive information (Bakshy and Messing 2015; Dubois and Blank 2018; Goel et al. 2010). Survey studies conducted across Europe, in Sweden (Dahlgren 2019), Spain (Masip et al. 2020), the Netherlands (Bos et al. 2016), and the UK (Fletcher et al. 2021b) have found limited evidence of the existence of echo chambers through self-selective exposure to partisan sources.

However, research has also found that citizens are rather passive with their information environments, as they have limited or no interest in controlling these (see Taber 2003; Ross-Arguedas et al. 2022). Studies have shown that individuals tend to mute, unfollow, and unfriend those who do not agree with them (Sibona and Walczak 2011). Moreover,research has found that exposure to information that contradicts one's opinion can have a boomerang effect of reinforcing their beliefs rather than reducing opinion differences (Bail et al. 2018; Jager and Amblard 2005; Macy et al. 2003). Hence, the concern about echo chambers and polarisation (Vaccari et al. 2016; Flaxman et al. 2016; Boutyline and Willer 2016).

Echo chambers and filter bubbles are two related but distinct concepts. While both phenomena involve the selective exposure to information that reinforces existing beliefs, they operate in slightly different ways and have distinct implications for political polarization and information consumption. Echo chambers are the result of self-selective exposure to certain content and people whereas the in-built platform algorithms (filter bubbles) are a product of personalized algorithms used by online platforms to tailor content recommendations for users. Both phenomena contribute to the fragmentation of public discourse and the exacerbation of political polarisation (Flaxman et al. 2016; Bakshy and Messing 2015; Guess et al. 2019). Yet, there are contradictory findings about the polarisation potential of the combined role of social media filter bubbles and online echo chambers (Ross-Arguedas et al. 2022; Dubois and Blank 2018).

Research on this topic has primarily focused on the presence or absence of echo chambers based of social media data (see Ross-Arguedas et al. 2022) while ignoring the importance of social networks for this process to occur. These online clusters are made up of connected individuals influencing each other. Opinion dynamics (OD) models have explored how the interactions between individuals, prompted by their attitude similarities (McPherson et al. 2001; Mäs et al. 2010) and social influence, can promote polarisation (Deffuant et al. 2000; Hegselmann and Krause 2002). Social simulation models have shown that the principle of bounded confidence is sufficient for attitude polarisation to emerge in a given population (Deffuant et al. 2000; Hegselmann and Krause 2002; Deffuant et al. 2002). These abstract models define opinion similarity as the difference between one's opinion and that of someone else. A pre-determined threshold bounds the agents' interactions and decisions to change their opinions (Amblard and Deffuant 2004; Jager and Amblard 2005).

Subsequent models explored the underlying social influence processes of opinion dynamics (Flache and Macy 2011; Flache 2018). Still, these types of models often have not included social networks into their simulations. The notable exceptions include: Flache and Macy 2011; Baumann et al. 2020; Banisch and Olbrich 2019; Milli 2021 which modelled opinion dynamics over artificial networks with various degrees of realism. Social networks composition, especially weak ties, or people who are not directly your close friends and family (Granovetter 1978), are crucial for getting exposed to cross-cutting content. Hence why it is important not just to focus on social media platforms and the information environment but also the social network composition when looking at polarisation.

Altogether, this paper proposes a theoretical synthesis of the political communication research assumptions about social media platforms and social influence along with the social simulation opinion dynamics models. It presents an agent-based model of social networks and the social media environment to explain the emergence of polarisation. This allows to unpack the mechanisms responsible for polarisation and observe under which conditions polarisation can emerge. This study focuses on national identity to provide a specific application of this model to a real-world problem such as inter-group conflict brought on by secessionist movements and the erosion of social cohesion through national identity polarisation.

Political opinions and partisanship have been the primary focus of this body of literature on polarisation (see Ross-Arguedas et al. 2022; Li and Zhao 2021; Barbera et al. 2015; Iyengar and Hahn 2009). There has been limited work on the issue domain of national identification despite of the direct implications for social cohesion (Holtug 2020; Richards 2013). The presence of secessionist movements brings out the question of national identity asking citizens to choose between them which may deepen the pre-existing divisions of national identity and prevent reconciliation (Elliott 2018). Previous studies have measured polarisation on a continuous scale ranging from 0 to 1 capturing the left and right ideological space (Fowler and Smirnov 2005; Singh et al. 2011; Deffuant et al. 2000; Schweighofer et al. 2020). This paper shifts the attention to national identity changes over time measured by the Linz-Moreno 5-point Likert scale question (Moreno 1995) on dual national identities. We conceptualise polarisation of national identity as the growing opposition of single versus dual national identities over time in a given population.

Our empirical case, selected to provide context for the simulation model, is Catalonia, the second most populated region of Spain. Since 2011, the Catalan secessionist movement has divided the Catalan and Spanish society on the basis of national identity. It has also shown signs of polarisation, reflected in the public opinion surveys collected by the Catalan Centre for Opinion Studies (*Centre d'Estudis d'Opinió*, CEO 2022). We use such data to inform the distribution of national identities in one version of the model to draw comparisons with an abstract random distribution, like simulation previous models (Deffuant et al. 2000; Hegselmann and Krause 2002; Macy et al. 2003). This enables us to explore the effects of one's social networks and social media's filter bubbles have on the resulting national identity dynamics.

As mentioned previously, social media platforms in-built systems promote selective exposure to certain content since they are designed to cater individual's tastes and likes as recommendation systems (Finkel et al. 2020). Some authors are sceptical to the idea of filter bubbles since they place more weight in the self-selective exposure than the algorithmic selectivity (Fletcher et al. 2021a). They argue that people can still choose to expose themselves to information that would contradict their views. Since determining the likelihood of people actively seeking cross-cutting content against the algorithmic default is out of the scope of this research, we propose to formally model two online social media environments. One where users are exposed to a more diverse diet and another where selective exposure to attitude-supportive content biases the media environment. This allows us to test the extent to which these in-built algorithms promote polarisation since, filter bubbles, in principle, can aggravate and may perpetuate the selective exposure-avoidance problem (see Nyhan et al. 2023 for recent review).

Political communication research confirms the existence of homophily on Twitter networks (Barberá 2014; Williams et al. 2015) and also on Facebook (Bakshy and Messing 2015). Our model integrates offline as well as online social networks to better reflect the duality of social networks as well as the effects of social influence. Both separate networks have their own topology to emulate the distinct features of offline social networks and online social media platforms. Moreover, we created two distinct network configuration scenarios to test the effects of homophily in one's network for the emergence of national identity polarisation. More details about this are provided under the model description section.

Overall, the purpose of the model presented here is to formalise social media platforms' filtering algorithms and the different social networks agents are embedded in to explain the emergence of polarisation. It goes beyond existing models to propose a different perspective on the subject by integrating theory while explicitly modelling online and offline social networks. It also provides a model of opinion dynamics applied to the Catalan secessionist movement with some empirically-informed parameters from that context to compare against an abstract model. Using this approach, enables us to untangle how social networks, social influence, and social media filter bubbles can contribute to the emergence of polarisation. In doing so, this work contributes to the existing literature of opinion dynamics and political communication by shedding light on the role of social networks and the media environment play in the polarisation processes through a theoretically-grounded simulation model.

The next section discusses the model specifications and assumptions based on the theories discussed thus far. This is followed by the results and discussion sections where the role of social networks and filter bubbles is explained and contextualised in the literature. Overall, we find that the initial network conditions in combination with the presence of filter bubbles are essential for the emergence of national identity polarisation.

Modelling social networks and social media filter bubbles

Agent-based models (ABM) are a computational method which allows to explicitly represent agents, their actions, and the environment in which these agent interactions take place (Gilbert 2008). Specifically, using an ABM allows to model the micro-level agent interactions and observe the resulting macro-level behaviours. This is not possible with other methods. ABMs have been used to model the COVID-19 pandemic (Almagor and Picascia 2020), organised crime (Elsenbroich et al. 2016) or civil violence (Epstein 2002).

Furthermore, these interactions tend to be non-linear, they cannot be added up to one another to explain the macro-level behaviours (Sayama 2020). The whole is more than the sum of its parts, to put it colloquially.

Polarisation is a collective phenomenon emerging from the actions and interactions of individuals. By studying the system as a whole, and representing individuals explicitly, we can be context-sensitive and understand better the mechanics responsible for its emergence. From a methodological perspective, we develop an ABM of national identity dynamics that enables us to test the extent to which social networks and social media platform filtering algorithms promote polarisation towards a single national identity versus a dual national identity. Overall, our model provides a possible and plausible explanation for how this process of polarisation can take place, applied to national identity.

General model description

Implemented in *NetLogo*, (Wilensky 1999), our model simulates a society (N = 2500) in which individuals, embedded in their offline and online social networks, interact with each other and the information they are exposed to. The model's population size of 2500 was chosen to provide a setting and to allow comparison between both versions of the model which start from a different distribution of national identities. The logic of the model goes as follows: each step individuals receive information from the media or from their social connections. Then, they decide whether or not to share it with their networks and then, they evaluate whether or not to update their national identities. The model and all data used can be found on Github and OSF. For more details see the Data availability and materials section.

Agents in the model

Agent-based models consist of three key elements: agents, attributes, and the environment in which the agents interact (Gilbert 2008). Agents represent individual entities modelled with the capacity to make decisions and interact with the environment they are in (Wilensky and Rand 2015). The agents, representing people in this model, have attributes responsible for their behaviours. The list of agent attributes can be found in Table 1.

The main attribute of interest in the model is national identification, a continuous variable [−1,1] inspired from the 5-point Likert-scale Linz-Moreno (Moreno 1995) dual national identity question. The categories were: "*I feel...*": *solely Catalan* (1), *more*

Table 1 Agent attributes, heterogeneous across agents (N = 2500) and drawn from random-normal distributions, $Z_i \sim N(0, 1)$

Parameter	Description	Ranges
My information	Represents the information received by the agent discussing national identity	[−1,1]
National identity	Indicates the agent's national identity and consequently group identity	[−1,1]
Uncertainty	Indicates the strength of an agent's convictions about their national identity	[0,1]
Engagement	Indicates the engagement levels or socialisation patterns of a given agent	[0,1]

Catalan than Spanish (0.5), *equally Catalan and Spanish* (0), *more Spanish than Catalan* (-0.5), *solely Spanish* (−1).

We initialised our model with two different national identity distributions. In one version, we used an abstract distribution ($\mu = 0$ and $\sigma = 0.40$), as opposed to a uniform one (Fränken and Pilditch 2018; Keijzer and Mäs 2022; Flache and Macy 2011) to avoid any overrepresentation of extreme opinions. In our second variant, we used survey data from the 2011 Catalan Centre for Opinion Studies (*Centre d'Estudis d'Opinión*, CEO 2022) on Catalonia's national identities. Fig 1 shows both distributions. We selected 2011 since it marked the start of the political institutionalisation of the Catalan secessionist movement as the Constitutional Court of Spain ruled regarding the Catalan Statute of

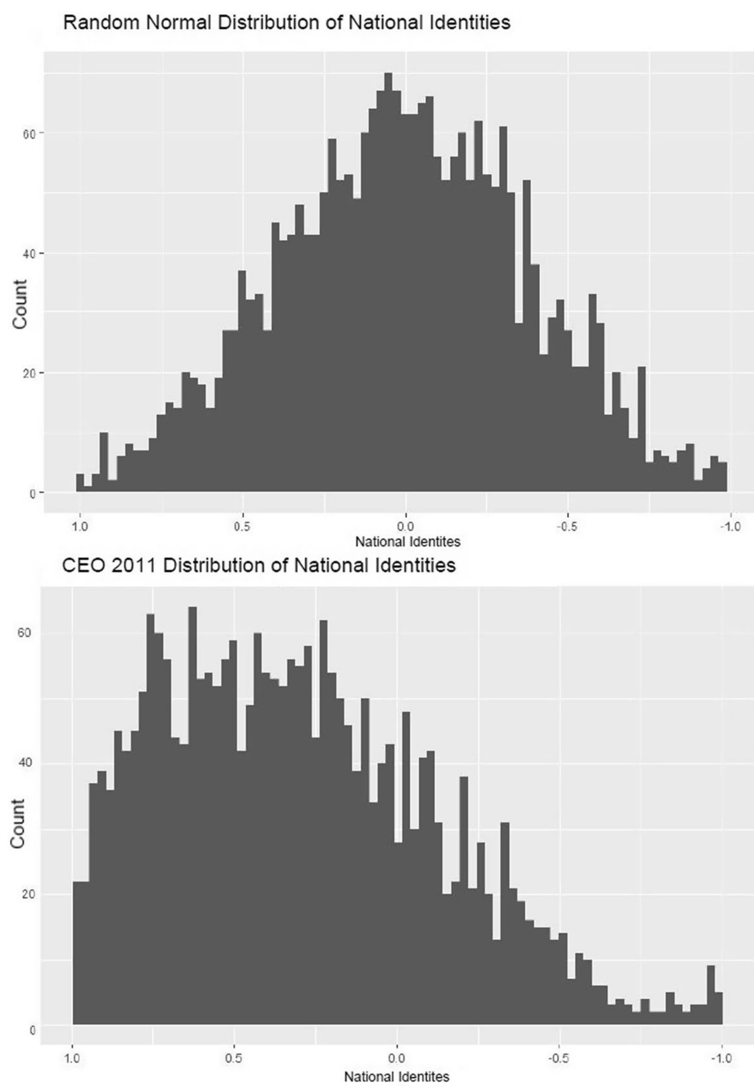

Fig. 1 Two distinct distributions of national identities in the simulation. A) Random normal distribution with $\overline{X} = 0$ and $\sigma = 0.40$ used for the abstract version. B) Right-skewed distribution from the CEO 2011 dataset with $\overline{X} = 0.45$ and $\sigma = 0.52$ used for the empirically-informed version of the simulation model

Autonomy and the political elites of Catalonia begun shifting in favour of secessionism (Lindez-Borrás 2013; Parella 2015).

Agents in the model have two other attributes besides their national identification. The *uncertainty* attribute, in Table 1, represents an agent's strength of beliefs and openness to accepting alternative ones. In this case, it represents their uncertainty about their national identities. It is randomly distributed (μ =0, σ =1) between 0 and 1 and heterogeneous across agents, unlike previous opinion dynamics models (Deffuant et al. 2000; Amblard and Deffuant 2004; Jager and Amblard 2005). The reason for making this an agent attribute and not a model fixed parameter is to acknowledge the susceptibility to persuasion or willingness to change one's mind (Gil de Zuniga et al. 2022) more realistically.

The *engagement* attribute, in Table 1, represents an agent's propensity to sharing an item of information received, the variable ranges from 0 to 1 and it is randomly distributed (μ =0, σ =1) across agents. This is independent of the national identity position in this model. It represents online activity patterns in an abstract manner to enable us to explore the interaction between social networks and filter bubbles. The motivation behind this parameter is to reflect people's varying degrees of online social media activities with some individuals being very active whereas others are lurking or less engaged (McClain 2021). The implications of formalising this assumption about engagement are discussed later on the paper. Moreover, we were not able to find data to empirically inform the uncertainty and online engagement parameters so we used a random normal distribution instead as a starting point. Future iterations of this model could explore the model's sensitivity to such distributions.

Lastly, the *my information* attribute, in Table 1, represents a piece of information an agent has received with a national identity value attached to it [−1,1]. The distribution of this parameter is the exact same as the national identity values (μ = 0 and σ= 0.40). Agents have limited memory. The value of my information gets updated each time they receive information from a tie or the media during a time step or tick. This stops when agents decide to update their national identity since they can only update their opinions once per tick. Moreover, agents do not have a memory of previous information received on previous steps, it is limited to the single tick or time step.

Agent's environment

The environment in which agents interact is a weighted multilayer multiplex social network, Fig 2, addressing the need to formalise the social networks agents are embedded in (Jager and Amblard 2005). These two layers are aiming to represent the hybrid media system with corresponding online and offline social networks of a given population. It is a multiplex network since the nodes on each layer represent the same entity, unlike interconnected networks, in which nodes in different layers represent different entities (Kinsley et al. 2020). Different layers in a multiplex network represent different relationships among the nodes (Boccaletti et al. 2014, p.17). In this case, the relationships are online and offline ties, which results in two distinct layers. The interlayer edges link nodes representing the same individual instead of different individuals across layers (Gómez et al. 2013).

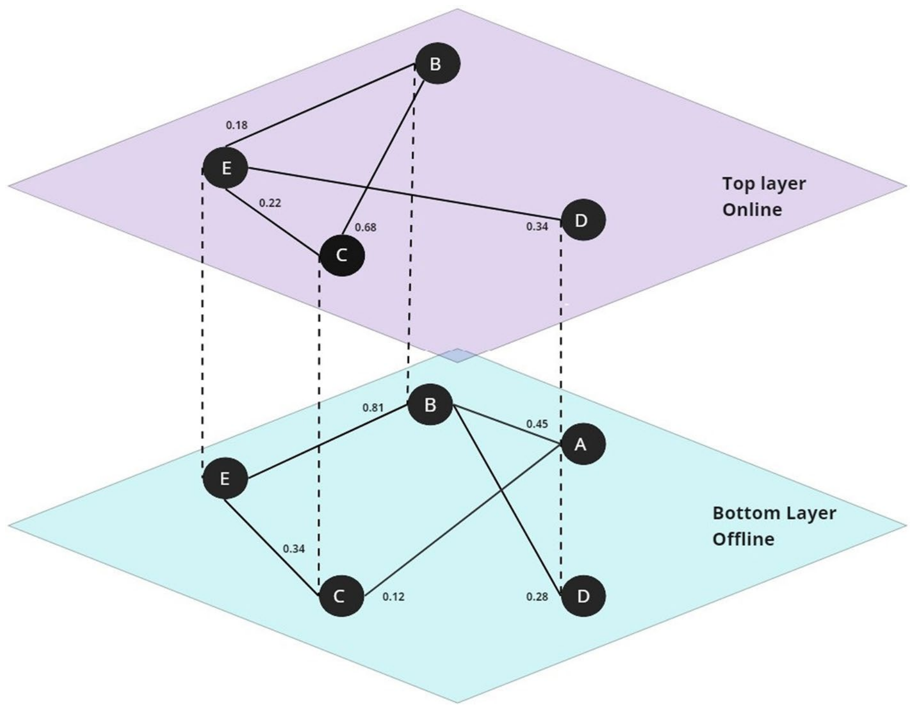

Fig. 2 The Bottom layer represents offline social networks and follows a Small-World (Watts and Strogatz 1998) topology. The Top layer represents online social networks and follows a Scale-Free topology (Albert and Barabási 2002). The nodes are the same across layers but not all nodes are present since not every person uses/has access to social media platforms. Full lines are intralayer connections between nodes and have weights assigned to them representing social influence. Broken lines are interlayer connections representing the same node across layers

The links or ties are weighted [0,1] in both layers or dimensions to represent the social influence capacity of each of the agents. Such weights are assigned at random and remain fixed throughout the simulation. Similarly, to the uncertainty and engagement agent parameters, these values were drawn from a random normal distribution ($\overline{X}$ = 0 and σ= 1) and have not been calibrated. Nonetheless, they represent the strength of different social relations. Lastly, the number of ties for each of the layers as well as the weights assigned to these remain fixed for the simulation.

The base layer represents an undirected offline social network and has a small-world Watts-Strogatz topology (Watts and Strogatz 1998) (n=2500, k=23, p=0.1). Agents have on average 23 ties or friendships and high clustering, characteristic of small-world networks. This value has been obtained from Lubbers et al. 2019 social network findings in Spain. The second layer represents an undirected online network on social media platforms and has a scale-free Barabási-Albert topology (Barabási 2009) (n=2000, k=45). The average node degree of the online networks is 45 which was scaled from Dunbar et al. 2015 findings about social media networks. This enables us to observe more clearly the effects of interpersonal social networks and the presence or absence of filter bubbles have on the emergence of polarisation. However, we recognise that such average degree values may produce a higher density of ties than we would observe in the real world which could potentially lead to a faster spread of information across the network. To test this, we ran additional simulations with a lower average degree (not included but available upon request) and found no difference in the resulting general patterns of polarisation, with the exception for the random networks in combination with filter bubbles where it slowed down the polarisation process.

One key feature of this model is not every agent has a social media presence. This is to reflect the fact that not every person uses social media platforms, in general but also to access the news (Gil de Zúñiga and Diehl 2017). The online to offline node ratio was obtained from Kemp 2022 . The proportion of social media users in Spain was of eighty percent so it was scaled down to the population size of this model which meant that 500 agents, out of the 2000 would not have an online presence in the second layer. This has direct implications as not every agent is exposed to the online filter bubbles, when active, echo chambers or is able to share information on the online networks.

Given these social networks, agents have two parallel communication regimes, inspired by (Keijzer 2022). Offline social networks in this model follow a one-to-one communication regime, meaning agents interact with one other agent each time. Conversely, online social networks follow a one-to-all communication regime like on social media platforms when someone posts on Facebook and all their friends will see it on their feeds. In practice this means that agents may receive multiple pieces of information from neighbours within a single time step or tick. They would process these sequentially and would stop doing so once they have updated their national identity, as they are only allowed to do so once per tick. These communication regimes have implications for the resulting national identity dynamics, as shown by (Keijzer and Mäs 2022).

The model environment has two independent network topologies for each of the layers with their characteristic structural properties. Additionally, we created two separate initial social network scenarios to test the effects of social network group composition in the polarisation process. In the Random initial social network scenario, agents are connected to each other following the topologies of the multiplex network environment as described previously and no rewiring promoting national identity similarity was performed on such networks to alter the group composition. In other words, the resulting social networks were diverse with groups of mixed national identities.

In a second scenario, the Homophily-based initial social network, agents are connected to each other on the basis of national identity similarity, following (Deffuant et al. 2000, 2002). The topology of the social networks remains a Watts-Strogatz and Barabasi-Albert, for the offline and online layers respectively. However, the edges on the online topology are rewired according to national identity similarity to enforce homophily of national identities. This scenario represents a situation in which agents are in homophilous clusters or echo chambers at the start of the simulation. This scenario promotes a majority of ties, over 90% of ties, within the similarity threshold. The similarity threshold, *s*, used here is the same as on previous bounded confidence opinion dynamics models (see Deffuant et al. 2000, 2002; Amblard and Deffuant 2004). We created this scenario of large majorities to compare against a random national identity mix in social networks and observe the resulting national identity dynamics. This scenario aims to capture the presence of online echo chambers where individuals are segregated into homophilous clusters.

Additionally, two social media environment scenarios were created to explore the role filter bubbles, generated by social media platform algorithms, play in the process

Table 2 Model parameters set at the start of each simulation and constant through time

Parameter	Description	Values
Information	Representing the media's reporting about the secessionist movement	N = 2500
Social networks	Two social network setup scenarios	Random; homophily-based
Social media	Two social media filtering scenarios	Filter bubble on; filter bubble off
s	National identity similarity threshold	0.5
d	National identity change constant	0.01

of national identity polarisation. These two scenarios only apply to agents part of the online social network since we are interested in reproducing the percolation of information from online to offline. The names of the random and selectively exposed scenarios correspond to filter bubble On and Off, on the model parameters in Table 2, for simplicity. The filter bubble Off scenario acts as a baseline scenario since here, agents have equal probabilities of getting exposed to national identity-supportive and national identity-discrepant information when they first receive information from the media. Having this scenario allows us to observe the evolution of national identities in a media environment where individuals get exposed to multiple sources of information, some that agree with them, others that challenge their national identities.

Conversely, the filter bubble On scenario simulates an unbalanced media diet whereby individuals get primarily exposed to information that supports their views (Iyengar and Hahn 2009; Sibona and Walczak 2011) or national identities in this model. In practice, agents that have online social networks would initially get information within their national identity similarity threshold s, but it could be closer to either bound of that information value, the upper or lower, based on a random probability. This scenario enables us to artificially test the effects of selective exposure to certain types of information by social media's sorting algorithms and the resulting effects on national identity.

The triple filter-bubble framework proposed by Geschke et al. 2019 formalised some of the longstanding assumptions about self-made and platform-imposed selective exposure to individuals and information. Yet, the social filter proposed by these authors was rather constraining and ignored the principles of social influence. The model presented in this paper, on the other hand, relaxes such constraint so that social influence prevails over one's national identity. Thus, maintaining the social connection between dissimilar agents, serving as a *weak tie* (Granovetter 1973), through which agents can get exposed to national identity-discrepant information. It should be noted that this process of information filtering is independent of the choice agents make in the model to share information which can either reduce or aggravate the selective exposure to certain kinds of information promoted by the filter bubble scenario. Altogether, having these four simulation scenarios allows to observe the effects social network composition and filter bubbles have on national identity polarisation and their interaction.

Table 2 shows all the model parameters, set at the start of the simulation run. The *Information* parameter represents an information pool of sources available to agents discussing national identities. These are generated at the start of the simulation and remain fixed throughout. They are drawn from a random-normal distribution ($\mu = 0$ and $\sigma =$

0.40). Similarly, the *Social Networks* parameter defines the initial social networks on the simulation, either a Random network or a Homophily-based network. Depending on the *Social Media* information regime selected, Filter bubble On or Filter bubble Off, the probability of agents getting exposed to certain content changes. This filter bubble On scenario only affects those agents present in the online social network layer. Beyond these, there are general model parameters, fixed throughout the simulation, that determine the national identity similarity threshold, s, and a constant, d, to normalise the rate of national identity change in the model.

Agent's interactions and decisions

The model presented in this paper aims to realistically represent the process of information receipt, evaluation, and sharing. Time in this model is treated as a discrete variable and it is abstract. This means that it advances one unit or step once all the agents have completed their actions or decisions, in line with previous simulation models (see Epstein 2013; Deffuant et al. 2000).

In this model, agents do not simply update their national identity following encounters with other agents, as seen in several previous models (Deffuant et al. 2000, 2005; Mäs et al. 2010). Here, agents have social networks that limit their exposure to information and other agents, as mentioned previously. This is to account for the social networks and hybrid media environment agents are embedded in. In doing so, agents in the model interact through their social networks instead of being bound by their opinion similarities like in (Hegselmann and Krause 2004; Deffuant et al. 2002; Flache and Mäs 2008). This allows us to explore the role social networks play in the polarisation process.

In every time step, agents receive information with a national identity value attached to it and have two independent decisions to make: information share and national identity change. First one being whether or not to share that information with their social networks. This decision depends on three factors: agent's engagement, one's social networks, and national identity similarity calculation.

Agents consider sharing information if their own engagement levels are greater than a random probability drawn from a normal distribution. Otherwise, agents would not be socially engaging with anyone and would move on to the next decision. This mechanism aims to represent different engagement levels found on social networks where some individuals may be more active, whereas others are rather passive or "lurking" (Rau et al. 2008; Rafaeli et al. 2004) without sharing among their networks. If their engagement levels are greater than such model parameter, they proceed to evaluate the national identity similarity. This means they calculate they distance between the information received and their own national identity, $NatSim_{ix} = abs(NatIs_i - NatIs_x)$. This evaluation follows the bounded confidence opinion dynamics models logic whereby every agent has a similarity threshold, $0 \leq s \geq |0.5|$ (Deffuant et al. 2005; Jager and Amblard 2005; Flache and Macy 2011). If the absolute difference between both values falls below the similarity threshold of 0.5, then the national identity overlap is sufficient to make the agent share the information with their ties. Sensitivity analyses were carried on this model parameter to justify its current value as the others were not a realistic evaluation. These can be found in the Appendix 1 for reference.

In line with the theories discussed in the previous section, individuals in the model get exposed to information through the media and their networks. While social media's filter bubble adds an additional layer to the selective exposure and selective avoidance processes in the model, an agent's social network acts as initial filter to information. This self-created selectivity integrates Noelle-Neumann's (Noelle-Neumann 1974) Spiral of Silence (SoS) cognitive process into the information sharing decision. The idea that individuals would refrain from sharing information that disagreed with the group's opinion in fear of being excluded due to their minority views (Chaudhry and Gruzd 2020). In doing so, this would limit the group's exposure to national identity-discrepant content. The model presented here incorporates that assumption as a threshold in the agent's decision to share information and tests the extent to which different initial social network configurations affect national identity polarisation while assuming SoS cognitive process is present in such context.

The SoS assumption becomes active when there's group consensus and the information received disagrees with such majority. Agents compute their online network's homophily coefficient. This calculation represents group conformity (Karimi et al. 2022) and measures the proportion of agents that share similar and dissimilar national identities to determine whether the agent belongs into the majority or the minority group in their online social network. An agent's network is considered homophilous if the national identities of over seventy percent of those tied to the agent are within the national identity similarity threshold, $s_{ij} \leq 0.5$. This model relaxes the majority and minority group proportions of Neuhauser et al. 2022.

Under these circumstances of group consensus and if the information's similarity calculation $NatSim_{ix}$ is greater than the similarity threshold, *s*, of 0.5 as mentioned previously, agents might only share with their social networks a small percentage of the time, 1/3 . In other words, when the information received exceeds the similarity threshold and there is group consensus in terms of a majoritarian national identity among agents, these would refrain from sharing such information most of the time but not always. This threshold value accounts for the fact that sometimes, people share information with their networks even if it contradicts the majority's opinions. We established this threshold as a reference point for this parameter since there is no data available to calibrate it.

Once the agent has decided to share the information they have received with their social networks, two parallel communication regimes become activated. A one-to-one communication regime is in place for the offline social network where one of the agent's offline ties gets chosen to receive information. This approach has previously implemented in Keijzer 2022; Keijzer and Mäs 2022. If the agent has online social networks, a one-to-all communication regime is in place where all the agent's online ties receive such information. This replicates the logic of posting on Facebook and all your friends seeing your post on their feed. These two regimes are very important for information cascades and national identity change since information can be coming from the media or from a tie. Agents may receive information from multiple of their social ties and evaluate each of them individually. However, once they have updated their national identity as a result of receiving information, they will stop evaluating incoming pieces of information during that time step.

The pseudocode of the decision to share information:

```
if Filter-Bubble On and any Online ties then
    Receive filtered (biased) info - more national identity supportive
     than -discrepant info
else
    Receive random info - same probability of national identity supportive
     & -discrepant info
end
while Agent Engagement ≥ random number do
    Compute national identity similarity Compute Group Homophily check
    if s ≥ 0.5 and there is Group Homophily then
        Share information - 20% of time (Spiral of Silence mechanism)
    else
        Share information
    end
end
Result: Share information
if Any Online ties then
    Share w/ all ties - One-to-All communication
     Ask ties to evaluate information
else
    Share w/ one tie at random - One-to-One communication
     Ask ties to evaluate information
end
```

The national identity change decision is the second decision agents make every step of the simulation. It expands on previous bounded confidence models (Hegselmann and Krause 2002; Flache and Macy 2011; Deffuant et al. 2005). After receiving information from the media or a tie agents evaluate such information on the basis of national identity similarity and compare it against their uncertainty parameter. An initial national identity was assigned to every agent, starting from a different distribution depending on the variant of the model. We formalise existing research suggesting the information from social connections tends to be more persuasive than from the media (Lazarsfeld et al. 1948; Weeks and Holbert 2013) in our model. If the agent receives information from a tie, the weight of that tie is added to the national identity similarity calculation, $NatSim_{ij}$. This is to reflect the social influence component in the decision to update the agent's national identity. In doing so, social influence is a contributing factor to the persuasion of the uncertainty parameter extending previous OD models (Flache et al. 2017; Amblard and Deffuant 2004; Deffuant et al. 2008) and making it more realistic.

The model presented here includes the three types of social influence, modelled by Flache 2018, since it is crucial for understanding the role of interpersonal connections play in polarisation. Assimilative influence (1), whereby agents' pre-existing national identity get reinforced with each interaction if the information they receive is within the national identity similarity distance threshold, $NatSim_{ij}$. It also allows for negative influence (2) as agents interactions are not limited by their national identity, like in other opinion dynamics models (Deffuant et al. 2005; Jager and Amblard 2005), but rather by the social networks they are embedded in. Hence the importance of having two social network scenarios that determine the initial network conditions for tie formation. Agents can therefore get exposed to national identity-discrepant information through weak ties (Granovetter 1973) and the media environment which could persuade them in favour of that national identity. The other type of social influence is (3) repulsive influence (Macy et al. 2003) when exposure to national identity-discrepant content will reinforce pre-existing beliefs meaning individuals will become more different to each other in terms of their national identities.

In the case that the national identity value exceeds the bounds set by the scale of the variable [−1,1], we truncate it to return it to the corresponding value within the distribution. The pseudocode of the decision to update their national identity:

```
if Information from tie then
    Compute national identity similarity
    if s + tie weight (social influence) ≤ (national identity) Uncertainty then
        Update national identity
    else
        Keep the same national identity
    end
else
    Information from the media if s ≤ agent uncertainty then
        Update national identity
    else
        Keep the same national identity
    end
end
Result: National identity update
```

Model results and analyses

Since the model developed in this paper used two different distributions of national identities, two sets of results and analyses are presented in this section. The first one corresponds to the abstract model version in which national identities were drawn from a random normal distribution. The second one set of results corresponds to the version of the model where the initial distribution of national identities in the population was based on the CEO 2011 data set on Catalan political attitudes (CEO 2022). A simulation run ends at t=400 since national identity dynamics become stable, as measured by no changes of the standard deviation not changing in the last 50 time following (Romero-Moreno et al. 2021). Other times were tested and no change was observed.

Table 3 shows the 4 configurations of social media and initial social networks for the two different initial distributions of national identity. The results were collected every 50 steps until the end of the simulation. These included the model set up parameters and agent attributes. In doing so, we were able to track each agent's changes in national identity over the course of the simulation. These individual changes are presented in Figs. 3 and 4 where each line on the graph corresponds to an agent. The population-level outcomes are presented later in the paper.

Table 3 Summary of the simulation configurations run in Netlogo BehaviourSpace and output measures for the eight testing scenarios

Social network scenario	Social media filter bubble	Nat. Id distribution
Random	Off	Random normal
Homophilous	On	CEO 2011

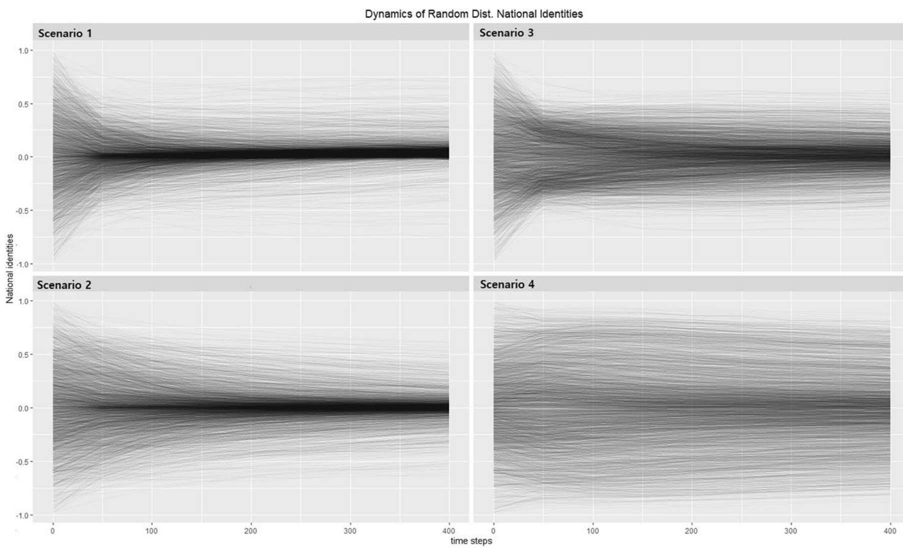

Fig. 3 Individual-level national identity dynamics starting from a random normal distribution of national identities over 400 model time steps. Scenario 1 = Random Network x Filter Bubble Off; Scenario 2 = Homophilous Network x Filter Bubble Off; Scenario 3 = Random Network x Filter Bubble On; Scenario 4 = Homophilous Network x Filter Bubble On

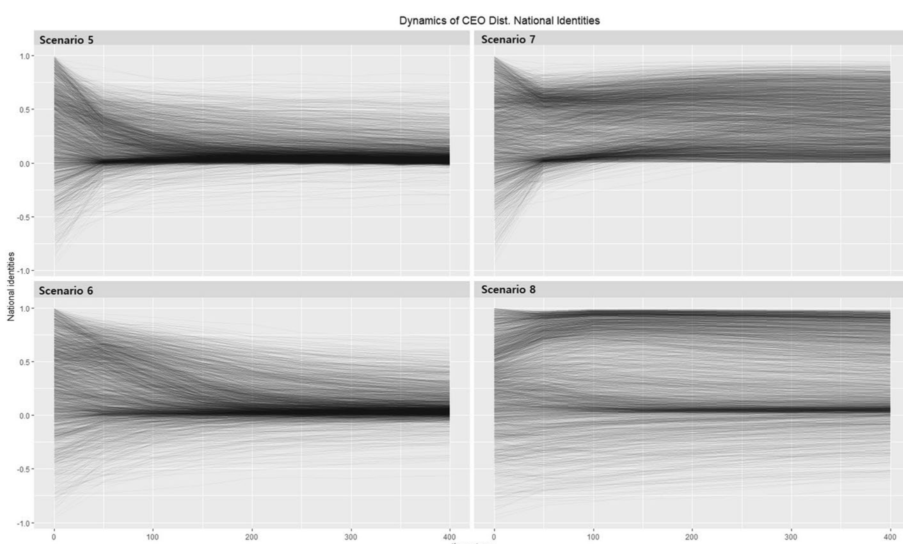

Fig. 4 Individual-level national identity dynamics starting from the CEO 2011 (CEO 2022) empirical-based distribution of national identities over 400 model time steps. Scenario 5 = Random Network x Filter Bubble Off; Scenario 6 = Homophilous Network x Filter Bubble Off; Scenario 7 = Random Network x Filter Bubble On; Scenario 8 = Homophilous Network x Filter Bubble On

Fig 3 shows the national identity dynamics in the first four scenarios which are the combinations of the *Social Network* and *Social Media* regimes in an idealised situation in which national identities are initially distributed randomly across the population. Various national identity convergence patterns can be observed across all four scenarios.

First, the social media filter bubble Off for both initial *Social Network* set ups, Scenarios 1 and 2 in Fig 3, produced national identity consensus, or single convergence (Deffuant et al. 2000), with the largest proportion of agents holding a national identity of zero, meaning *feeling equally Catalan and Spanish*, and close to no agents at either end of the national identity scale. Yet, there is a difference in the rate of convergence between the social network initial setup whereby random networks, where ties are made regardless of national identity similarity. National identities converge faster and are more stable compared to homophilous networks, in the absence of filter bubbles.

The presence of social media filtering algorithms for the two *Social Network* scenarios, Random and Homophilous, scenarios 3 and 4 in Fig 3, produced moderate convergence of national identities. Especially, the combination of filter bubbles and an initial homophilous social network setup (scenario 4) preserves most of the initial variation in national identities. It produces only moderate convergence at the centre. Alternatively, the random social network scenario, combined with the selective exposure of filter bubbles (scenario 3), produced a moderate convergence towards the centre, but not the single national identity cluster which emerged in the absence of filter bubbles.

Fig 4 shows the national identity dynamics in the last four scenarios of Table 4, in which the initial distribution of national identity was based on the CEO 2011 survey data set (CEO 2022). In these model scenarios we observe similar national identity dynamics to those in Fig 3, with a constant clustering around zero *feeling equally Spanish and Catalan* and varying proportions of agents at either extreme of the national identity spectrum. This suggests that the national identity dynamics observed in the random distribution model version are well reproduced in the model version initialised with the rightly-skewed distribution of the CEO 2011 survey data.

Similar to the other four scenarios, the effects of the social media filter bubble were limited in the random social network setup, scenarios 5 and 7 in Fig 4, in terms of producing national identity polarisation for this version of the simulation. Single convergence towards zero, representing *feeling equally Catalan and Spanish*, can be observed on both scenarios but at different rates. Initially homophilous social networks and no filter bubbles, scenario 6, converge slowly as more agents can be found in the upper bound of the distribution, between 0.5 and 0.75 meaning *feeling more Catalan than Spanish* or

Table 4 Gini coefficient of the distribution of national identities for each of the four scenario combinations across both versions of the model, abstract and CEO 2011 collected using NetLogo BehaviourSpace

Social media	Abstract dist.		Empirical dist.	
	Social networks			
	Random	Homophilous	Random	Homophilous
Filter On	0.12	0.21	0.11	0.17
Filter Off	0.08	0.11	0.09	0.12

solely Catalan. This is not as prevalent in the initially random social network scenario (5) where we see a faster convergence from a single national identity to a dual national identity around zero. This suggests that homophily within social networks, a situation which reproduces the phenomenon of echo chambers, does not prevent convergence towards a single national identity cluster, but reduces the pace at which convergence is produced when the media environment is mixed (filter bubble off).

On the other hand, the scenarios in which selective exposure to national identity-supportive information is enabled by the filter bubbles, scenarios 7 and 8 in Fig 3, produce two different national identity dynamics depending on the initial social network condition. The homophilous social network setup in the presence of filter bubbles, scenario 8, produces national identity clustering and a bi-polarisation dynamic. This is the only scenario where the proportion of agents at the lower bound of the distribution, -0.5 to −1, representing *feeling solely Spanish* does not disappear but rather remains constant over time. Conversely, this pattern is less clear on the random social network scenario 7 in Fig. 3, where agents converge into two parallel clusters from the extremes of the national identity distribution early in the simulation while also maintaining a relatively high number of agents in-between both national identity clusters.

Beyond the individual-level modelling outputs, we also collected population-level national identity polarisation data in the same 50-step intervals as the individual-level data. Previous studies have used the variance of a distribution as a measure of polarisation (Lelkes 2016; Deffuant et al. 2002; Pardos-Prado and Dinas 2010), in our case we use the Gini coefficient since it is a more robust measure of the shape of the distribution (Badham 2013). This measure ranges from 0 to 1 to indicate perfect equality and inequality in a distribution. Given that our outcome variable range includes negative values, we adjusted the scale by adding 1 to all values before calculating the Gini coefficient. Table 4 presents such results for both versions of the model, averaged over a hundred repetitions of a given parameter combination.

The first thing we observe is that the Gini coefficient is moderately low across all simulation scenarios meaning moderate equality across the distribution of national identities for both versions of the model. Yet, we can observe that initial social networks where individuals were in a diverse group (Random in the parameter configuration) produced less inequalities or polarisation compared to those where the initial social network that promoted group similarity (Homophilous in the parameter configuration). Scenarios where social media filter bubbles were present, regardless of the distribution of national identities in the population (Fig 1), had the largest Gini coefficient values ranging from 0.11 to 0.21. In particular, scenarios where the initial social network was homophilous and agents were selectively exposed to primarily supportive information through filter bubbles produced the greatest largest Gini coefficient, Tab 4. This suggests more inequality in the distribution of national identities or that polarisation may be present, as the individual-level data has shown (see Figs. 3 and 4).

Moreover, those simulations that initialised the agent's national identities from the CEO 2011 data set produced slightly larger values of the Gini coefficient for all scenarios, compared to those from the random distribution. Meaning that starting the simulation from this empirical distribution and exploring the different scenario combinations resulted in more unequal distributions of national identities across the population. The only exception to this was observed for the scenario where filter bubbles were present in combination with an initially homophilous social network. Under such conditions the Gini coefficient was 0.17 compared to 0.21 of the abstract random distribution. Nonetheless, looking at the individual-level findings from 4, scenario 8, we observe more clustering around the upper bound of national identity corresponding to *feeling solely Catalan* and around the centre value zero corresponding to *feeling equally Catalan and Spanish* and fewer individuals in-between both values suggesting the presence of polarisation but not the expected bi-modal distribution as shown in (Deffuant 2006; Mäs and Flache 2013). These patterns are less clear in the abstract model outcome scenario 4 in Fig. 3 with more spread out values across the whole scale and slightly higher Gini coefficient. Overall, this section has presented the results from the two independent model set ups and the effects of social media and social network scenario combinations had on the resulting national identities at both the individual- and the population-level.

Discussion

This work has developed a theoretical agent-based model to study the effects social networks and social media filter bubbles in the polarisation process. The simplicity of the model highlights their effects both individually and in combination. This model has been applied to the case study of Catalonia's national identities in the context of the ongoing secessionist movement. Data from the Catalan Centre of Opinion Studies (CEO 2022) has been used to empirically-inform the distribution of national identities in one version of the model. The agents in this model are connected through social networks on two layers representing offline network and their online social media counterparts. Moreover, they interact not just within their networks but also with a hybrid media environment with varying exposures to selectivity bias of information and people. Therefore, formalising existing political communication and social influence theories and empirical findings about the role social media platforms and social networks play in the process of polarisation.

Research has shown that the presence of a social connection between the sender and the information receiver directly affects the persuasion of the information received (Metzger and Flanagan 2013; Metzger et al. 2010). Individuals modify their believes and attitudes in an attempt to resemble those with whom they interact with (Axelrod 1997). Lazarsfeld and colleagues (Lazarsfeld et al. 1948) found that is person-to-person influence has the most effect on those susceptible to change their opinion. Three types of social influence have been identified in the literature (Flache et al. 2017): assimilative influence, similarity based, and repulsive influence. The first type reflects the earliest models of social influence where influence is bi-directional, reducing opinion differences between people (Festinger 1957). In other words, it represents the rate of opinion convergence, how long it takes for a group to converge around one opinion (Flache et al. 2017).

Previous opinion dynamics models have shown central convergence of opinions where extremists are unable to exert influence on the population at the centre of the opinion distribution (Deffuant et al. 2002). In this paper's findings, we also observe this same convergence towards the centre, representing *feeling equally Catalan and Spanish* on the Linz-Moreno scale (Moreno 1995). This happens where an agent's initial social network is composed of national identity-supportive and national identity-discrepant others with equal likelihood. This exposes agents to diverse national identities in both models, abstract and empirically-informed (see Figs. 3 and 4) where assimilative influence promotes national identity convergence. This is particularly relevant for the model version which used the CEO 2011 data set of Catalonia's national identity distribution. It showed how in the absence of selective exposure to primarily national identity-supportive information through filter bubbles and through assimilative social influence from social networks, the originally polarised distribution moderates over time towards consensus around a dual national identity *feeling equally Catalan and Spanish*. Perceptions of majority and minorities in social networks, as evidenced in Karimi et al. 2022 and theorised by Festinger 1957; Latané 1996, help us explain these patterns. When individuals find themselves holding a minority view in a cohesive social group, they will be persuaded to change their views to conform to the group's, therefore creating consensus. We observed similar patterns in our results about national identities (see Figs. 3 and 4).

Nonetheless, we also observe convergence of national identity towards a single national identity cluster, at the centre of the distribution, corresponding to *feeling equally Catalan and Spanish*, in the initial homophilous social network in the absence of filter bubbles. This occurs at much slower rate without reaching a single cohesive national identity cluster like in the initially random social network scenario (see Figs. 3 and 4). Granovetter's work (Granovetter 1973) is relevant in this regard since it exemplifies that it is through *weak ties* or social connections individuals get exposed to alternative information and perspectives. These weak ties represent individuals whose national identity differences are moderate but not sufficiently large to be too dissimilar and break off that social connection. The strength of the social connection explains why those weak ties are maintained over time. Having such connections allows for exposure to national identity-discrepant information despite of similarity bias social influence being present.

We observe the strength of weak ties too where the initially homophilous networks, with a majority of national identity-supportive ties, in the absence of social media filter bubbles tend to converge towards zero forming a single large national identity cluster. Under these conditions, initially extreme individuals that felt *solely Catalan* or *solely Spanish*, under Linz-Moreno terms (Moreno 1995), shift their national identities in favour of the majority as they get exposed to diverse information and diverse views evidencing the importance of weak ties for slowing down the polarisation of national identity. These results are in line with political scholars who found that a diverse media diet, getting exposed to national identity-consistent and national identity-disagreeing information, reduces opinion differences and promotes consensus (Balietti et al. 2021).

Similarly, a new wave in computational social science has proposed bridging algorithms on social media platforms as a way of increasing mutual understanding by promoting content that will promote agreement across groups that hold opposite views (Ovadya and Thorburn 2023). While our paper did not formally use such algorithms, the filter bubble Off scenario did offer a more balanced information diet that showed a shift towards a single opinion cluster under both initial social network scenarios instead of increased polarisation. Future works could explore implementing different filter bubble algorithms and observing their effects, in combination with different social networks, on polarisation.

Yet, previous research on social media platforms (Bail et al. 2018) have found that when partisans get exposed to a diverse media diet this promotes national identity reinforcement, or repulsive influence in social simulation terms (Macy et al. 2003; Jager and Amblard 2005). A recently published study about the effects of social media's algorithmic filtering on opinion polarisation (Nyhan et al. 2023) partially explains this slower shift towards a single national identity we observed in the simulation scenarios where filter bubbles were absent and the initial social networks were homophilous (see Scenarios 3 and 6 in Figs. 3 and 4). Their findings suggest that even when opinion-supportive information is decreased in one's Facebook feed, this would not in turn decrease polarisation as it was previously found (Guess 2021).

Our findings suggest that repulsive influence may not be as strong as those studies claim. This is especially the case if we take into account the social networks individuals are embedded in as that moderates the polarisation process in the absence of filter bubbles. This was the case when the initial social networks where homophilous as we altered the group composition towards a majority of like-minded agents as opposed to a diverse group of national identities. The strength of repulsive influence was indirectly affected by this as shown in the individual- and population-level modelling outcomes (see Figs. 3 and 4 and Table 3). In the absence of social media filter bubbles, even those agents at the extremes of the national identity distribution, *feeling solely Catalan* and *feeling solely Spanish*, tend to converge towards the centre as they are a minority that is getting exposed to diverse others and information. This intermediate homophily range produced by random networks promotes minority nodes to converge to the majority views, as demonstrated by Karimi et al. 2022. Congruent with those findings are the results presented here where individuals get exposed to a diverse media diet, as filter bubbles are absent, those individuals on the extremes tend to shift their national identities in favour of the moderate majority holding a dual national identity *feeling equally Catalan and Spanish* (see Figs. 3 and 4).

Moreover, very limited number of studies has explored the effects of filter bubbles on opinion change from a simulation perspective, with a notable exception (Geschke et al. 2019), which formalised some of the longstanding assumptions about self-made and platform-imposed selective exposure to individuals and information. Nonetheless, their social filter was rather constraining and ignored the principles of social influence and kinship. The model developed in this paper, on the other hand, relaxes such constraint so that social influence prevails over one's national identity, thus maintaining the connection, serving as a *weak tie* (Granovetter 1973) through which individuals can get exposed to opinion-discrepant information. Furthermore, this model formalised two independent scenarios of social media filter bubbles to isolate the impact that these inbuilt platform algorithms, selectively exposing individuals to primarily opinion-supportive content, have on the overall process of opinion change. Therefore, contributing to understanding the role these filter bubbles have and also allowing to quantify the effects selective exposure, through filter bubbles, has on opinion dynamics which has been a longstanding demand in the political science literature (Ross-Arguedas et al. 2022).

Beyond the artificial selective exposure of social media platforms, information diffusion plays a crucial role in polarisation and *othering* processes, given the growing importance of social networking sites for news sharing (Kumpel et al. 2015). Differing patterns of political opinion diffusion online have been found depending on the group size and homophily levels of the social network (Halberstam and Knight 2014). Politically engaged individuals tended to have greater networks, with more connections, and thus more information supporting their political opinions. Moreover, Bakshy and colleagues' Facebook experiment demonstrated that tie strength was directly impacting the diffusion of information (Bakshy and Messing 2015). In particular, weak ties consume and transmit information that one is unlikely to be exposed to, thus increasing the diversity of sources of information on a given network. These findings were also found on Twitter (Arnaboldi et al. 2016). This provides further support for creating two separate initial social network environments that varied the group composition in this paper's ABM model. Thus, allowing to observe the direct effect of social networks information diffusion have on national identity polarisation. Under the random initial social network and in the presence of social media filter bubbles, the higher density of network ties speeds up the process of polarisation due to the higher information diffusion rate promoted by mixed group composition and biased media environment. In contrast, initially homophilous networks are constrained by the majority and minority perceptions gate-keeping the information to be shared beyond the number of neighbours. Additional experimental results on network density are available upon request. As a result, a clearer picture of the combined role of social networks and social media's filtering algorithms can be drawn. This is much need in the literature since the extent to which echo chambers exist on social media and increase online segregation has been contested (Dubois and Blank 2018; Cioroianu et al. 2018; Goel et al. 2010).

Limited empirical evidence has been found of the existence of echo chambers through self-selective exposure to opinion-supportive content and its promotion of polarisation (Masip et al. 2020; Fletcher et al. 2021a; Tucker et al. 2018). A social simulation approach like the one offered in this paper is capable of addressing this gap in the literature regarding the extent to which social media platforms' filtering algorithms and social networks contribute to the emergence of echo chambers, and polarisation more broadly. Similarity bias models of social influence (Deffuant et al. 2000; Hegselmann and Krause 2002; Macy et al. 2003) are relevant in the context of echo chambers since their core assumption is that interactions between individuals are prompted by their opinion similarities (McPherson et al. 2001; Mäs et al. 2010). Political communication research confirms that birds of a feather do flock together on Twitter networks (Barberá 2014; Williams et al. 2015) and also on Facebook (Bakshy and Messing 2015). This paper has shown that starting from a homophilous social network where online echo chambers were present, increased in the polarisation of national identity compared to initially random networks. Therefore suggesting that indeed, similarity does breed similarity in social networks. The social simulation approach provided in this paper was able to show how this phenomenon may take place. This is in line with previous findings on the negative impacts of echo chambers for polarisation (Barbera et al. 2015; Guess et al. 2020; Williams et al. 2015). Thus, adding further evidence suggesting that echo chambers' effect on online segregation maybe greater than expected (Flaxman et al. 2016; Dubois and Blank 2018; Garrett et al. 2013).

The combination of social media filter bubbles with echo chambers has been an ongoing concern in the politics literature (Ross-Arguedas et al. 2022; Fletcher et al. 2021a;

Cardenal et al. 2019; Zuiderveen Borgesius et al. 2016) due to its polarisation potential or lack thereof. This paper tested interaction between social media filter bubbles and echo chambers for both model versions, abstract and empirical, in scenarios where the filter bubble was On and the initial social networks were homophilous. When looking at the polarisation of such social networks the picture is rather different in the presence of social media filtering algorithms as shown by Figs. 3 and 4. The average proportion of agents on the extremes of national identities is greater but it is only when combined with a homophilous social network setup that the bi-polarisation pattern truly emerges, like on Deffuant et al. 2000, 2005; Flache 2018 but not as extreme two-national identity clusters and no individuals in-between as previous opinion dynamics models have found.

Furthermore, negative social influence can help explain this phenomenon since here, individuals are not only exposed to like-minded others through their social network configuration but also through the filtering algorithms on social media that reinforce their national identities (Mäs and Flache 2013). This is particularly evident among individuals in the middle of the distribution those that felt slightly *more Catalan than Spanish* which are drawn towards either extreme, feeling *solely Catalan* or towards *feeling equally Catalan and Spanish* due to their social networks social influence and the prevalence of selective exposure through filter bubbles.

In the case of random networks, the explanation is slightly more complex since here individuals have diverse social networks while being exposed to information that supports their existing national identity through social media filter bubbles. Exposure to information that disagrees with one's views has been found to reinforce existing beliefs (Macy et al. 2003; Baldassarri and Bearman 2007; Del Vicario et al. 2016) instead of promoting the opposite opinions. We would then expect then that individuals in an initially random social network, would become more polarised over time due to this repulsive social influence as they are exposed to national identity-supportive and national identity-discrepant others. Yet, we see that in the presence of social media's filter bubbles, promoting national identity-supportive information that not entirely the case. Once more highlighting the effect of social media features like filtering algorithms in promoting polarising content despite of individuals not setting out to get exposed to such content, as shown by (Settle 2018). Yet, when these differences of opinions are too large it may actually promote consensus (Takacs et al. 2016) or less defined national identity clusters as seen in the random social network and social media filter bubble scenarios. These simulation results bear resemblance to real-life context where individuals may be segregating themselves into national identity clusters but yet, polarisation requires of social influence to emerge (Flache and Macy 2011) as shown in the homophilous social network and social media filter bubble condition.

Lastly, the presence of echo chambers and social media's selective exposure through filter bubbles was shown to increase the polarisation of national identity in both model versions initialised with a different national identity distribution. These findings are in line with much of the political communication literature suggesting the existence of echo chambers on social media (Tucker et al. 2018; Fränken and Pilditch 2018; Geschke et al. 2019) and provide further support for the prevalence of echo chambers disputed in the literature (Dubois and Blank 2018). Therefore providing empirical support for the effect of both, online echo chambers and filter bubbles, in the promotion of national identity polarisation. Overall, our model and findings highlight the importance of social network configurations, social influence, and filter bubbles in the polarisation of national identity.

However, this research has some limitations and weaknesses. First, we used the Catalan secessionist movement to test the mechanisms responsible for the emergence of polarisation, applied to national identities in this case. Using a single case study is often problematic as it limits the generalisability of the findings (Seawright and Gerring 2008) as no other cases were used to compare against. Future research could explore if the same mechanisms apply in different contexts.

Second, the value chosen to represent the assumption from the Spiral of Silence theory (Noelle-Neumann 1974; Gibson and Sutherland 2020) was homogeneous across agents, constant throughout the simulation and not drawn from empirical data. This value represented the likelihood that an agent would share national identity-discrepant information with their social network when there is a majority-view in the group. This issue could be addressed in further sensitivity analyses to determine whether such value has an effect on the resulting national identity dynamics, but it is outside of the scope of this paper.

In a similar fashion, greater online engagement has been found to be correlated with holding more partisan or extremist views as well as higher exposure to misinformation (Chen et al. 2021). This can in turn increase polarisation (Barbera et al. 2015). Yet, the model presented here assumes online engagement is independent of holding extreme views as it interested in exploring how interpersonal social networks and online filter bubbles can foster polarisation, not the effects of online engagement on polarisation itself. Future studies could look at how online engagement can affect polarisation and the interaction of such relationship with one's social network and social media's filter bubbles in this context.

Lastly, we acknowledge that the average degree used to inform our social networks may produce greater density of ties than in the real world and therefore faster spread of information. Yet, our goal was to formalise these online and offline networks and compare their effects on polarisation in combination with two different social media environment scenarios. Also, the homophilous social network scenario may have implications for the underlying structure of the online network layer. We used the initial network topology of Barabási-Albert and rewired based on national identity similarities which could have altered the structural properties of the network. Future work aims to explore the differences in path lengths and centrality measures to see whether these are essential for the national identity dynamics observed in the model.

Conclusion

To summarise, this paper aimed to explain the role social networks and social media filtering algorithms play in national identity dynamics in a context where secessionist movements were present. From a political communication perspective, this paper combined the existing approaches of opinion dynamics and social influence. We developed an agent-based simulation with two distinct and interrelated mechanisms, social networks and social media filter bubbles. The Catalan secessionist movement was chosen to provide context for this model and survey data from the Catalan Centre of Opinion Studies (CEO 2022) was used to empirically-inform the distribution of national identities. This allowed for comparison between the abstract theoretical model with an initial random distribution of national identities and the CEO 2011 version of the model and exploring the effects of the various mechanisms at play.

The model presented here makes two contributions to the existing social simulation and political communication literatures. First, it explicitly models social networks in the form of a weighted multilayer multiplex network of online and offline social networks. Limited research has been explicitly modelled social networks in opinion dynamics models, with some exceptions (see Geschke et al. 2019; Macy et al. 2003; Flache and Macy 2011). As shown by the results, social networks were a crucial component to understanding the emergence of national identity polarisation. The second contribution this paper makes is that it shows how the presence of social media filter bubbles selectively exposing agents to national identity-supportive information promotes multiple national identity clusters and national identity polarisation under different initial social network compositions.

Overall, this paper has provided valuable insights into the role social networks and social media filter bubbles play on the emergence of polarisation through the national identity dynamics simulation model developed. The findings suggest that while social media filtering algorithms may promote certain content for its users, social network configurations are ultimately essential for polarisation dynamics to emerge. While abstractly representing the Catalan context, the simulation model presented here was able to show how through social media filtering algorithms and one's social network national identity polarisation may take place. Therefore, offering a possible and plausible explanation for how such polarisation may have taken place.

Appendix

The bounded confidence threshold of national identity similarity, s, has been used in previous opinion dynamics *OD* models with different values and distributions across agents (see Amblard and Deffuant 2004; Jager and Amblard 2005; Mäs et al. 2010; Deffuant 2006). This parameter is an abstract value that cannot be derived from data since it represents a person's tolerance or openness to other's opinions. Sensitivity analyses were conducted on s to test the extent to which national identity dynamics are sensitive to it. This parameter is present in three separate instances of the model. The first one being used to build the homophilous social network. When agents are deciding who to befriend at the start of the simulation when the social networks are created, they evaluate the national identity differences, between themselves i and their neighbour j. If such difference is below s, then the neighbour is deemed similar.

The other two instances in which this parameter, s, is relevant is information sharing and national identity change decisions. The information sharing decision involves computing one's neighbours' national identity similarities, relevant for the group categorisation process. Furthermore, the bounded confidence threshold of national identity similarity *(s)* is used to determine whether the information an agent receives supports or disagrees with the agent's national identity. In turn, this affects the national identity decision.

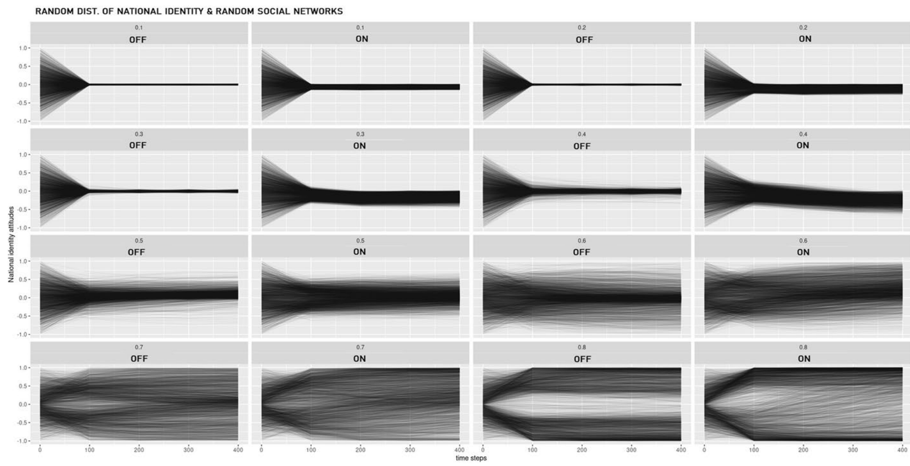

Fig. 5 Sensitivity Analyses of *s* for initial Random social network and the two Filter Bubble conditions

For both Social Network and Social Media Scenarios, *s* values between 0.1-0.8 were tested. Ten runs were made for each parameter combination and they were averaged across runs (N= 160). Only the abstract random distribution version of the model as that is the original model we compare against an empirically-informed version. After testing for the same parameter combinations, the bounded confidence threshold of national identity similarity (*s*) was kept at 0.5 for the initial homophilous social network set up, which establishes the initial social networks based on national identity similarity, as no noticeable interaction effects were discovered in the resulting national identity dynamics. This means that the information sharing and national identity change agent decisions are sensitive to this parameter *s*, under the initial social networks and social media filter bubble conditions.

Figure 5 shows the various values of *s* for initial Random social networks in the presence or absence of Social Media Filter Bubble conditions. The lowest value of *s* represents instances where the absolute difference between the agent's national identity and the information received or their neighbours' national identity, as this parameter is used to calculate both, is very small. When *s* is below 0.4 values, irrespective of the Social Media filter bubble condition, we observe single converge around zero or feeling equally Catalan and Spanish. There is a larger proportion of agents towards the centre of the distribution, since the distribution of national identities in the population is normally distributed. This draws in other agents from the extremes due to the small bounded confidence national identity similarity threshold *s*. As a result, all agents adopt a dual national identity of feeling equally Catalan and Spanish, forming a sole cluster sustained over time, regardless of the Social Media filter bubble condition.

We can also observe that as this parameter increases, there is more fluctuation around the centre of the distribution where the main national identity cluster is located. These fluctuations are more visible in scenarios where the filter bubbles are present (second and fourth columns) in Fig. 5. This can be explained by the fact that these filtering algorithms selectively expose individuals to national identity supportive information which in combination with the initial social networks produces more

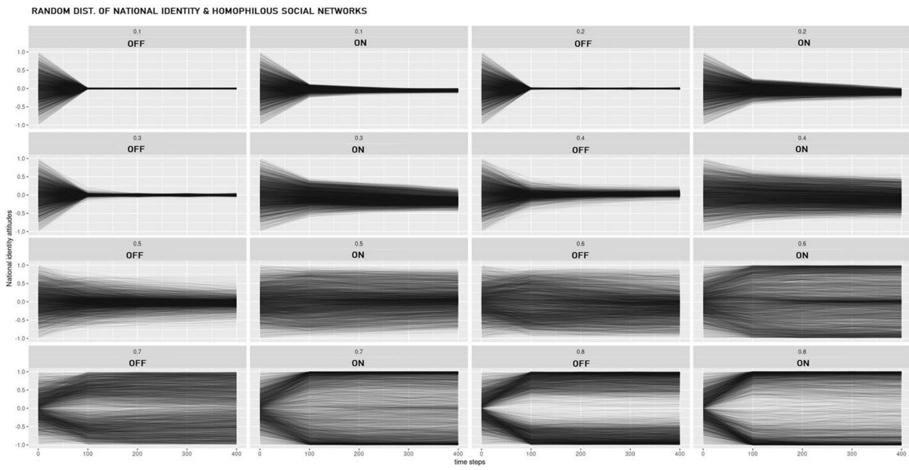

Fig. 6 Sensitivity Analyses of *s* for initial Homophilous social network and the two Filter Bubble conditions

variability of national identities. There is a tipping point where the single national identity cluster becomes two parallel ones, between 0.7 and 0.8 values of *s*. It is in these two scenarios that we observe the traditional polarisation dynamics of previous opinion dynamics models (Flache and Mäs 2008; Amblard and Deffuant 2004). Agents have, still, a diverse social network with agents of various national identities since they were initialised from a random social network where ties were created, at random regardless of national identity similarity.

We can observe that the greater the value of *s*, or larger national identity similarity threshold, the greater the number of national identity clusters. We can see that in the presence of filter bubbles where agents mainly would get information within their national identity similarity threshold, *s*, convergence towards a single national identity cluster is reduced as *s* increases. Under these conditions, there is more scope for diverse views, especially when *s* is but this does not necessarily mean that agents get persuaded since their (national identity) uncertainty is heterogeneous and unique to each agent. Moreover, these scenarios where *s* is greater than 0.5 are not realistic because individuals may interact with others that completely disagree with their own national identities, at the other end of the national identity distribution, but this would not bring them any close as they are very different. Instead, this will act as a repulsive force that would promote extremist views instead of convergence, as shown by repulsive models of social influence (Flache et al. 2017).

A rather similar picture can be found when the initial social network is homophilous instead of random, see Fig. 6. Yet, there are slight differences between the resulting national identity dynamics. We observe that lower values of national identity similarity threshold, *s*, produce national identity convergence towards zero. In the presence of social media filter bubbles this convergence is slower as we observe more agents diverging from the centre of the distribution towards the values of (+/-0.5) and eventually reaching the extremes of the distribution at the highest value of *s* I tested. Under those conditions, the higher values of *s* produce a clearer bi-polarisation pattern, especially 0.7 and 0.8 values, unlike in the initially random social network scenarios. Here, individuals were getting exposed to information that reinforced such views, through their homophilous social networks, increasing the proportion of agents at the upper and lower bounds of the national identity distribution [1,−1].

Yet, when the value of the bounded confidence threshold of national identity similarity *s* is 0.7 and 0.8 (see Fig. 6) we observe that the absence of filter bubbles moderates the emergence of polarisation which is expected since agents will get exposed to diverse information while being in a homophilous social network with a majority of neighbours that have a similar national identity. This means that agents will maintain their national identity instead of changing it in favour for the more extreme one. However, as this national identity difference becomes greater, *s* being greater than 0.6, these differences between agents only grow in favour of the extremes due to the large tolerance to different national identities.

As mentioned at the start of the sensitivity analyses, the mechanism responsible for creating the homophilous social networks was set to 0.5, in terms of the bounded confidence threshold of national identity similarity. This means that initially those agents at the centre of the distribution, *feeling equally Catalan and Spanish* (0), get exposed to either neighbours *feeling more Catalan than Spanish* (+0.5) or *feeling more Spanish than Catalan* (-0.5) since this national identity similarity remains at 0.5. This means that as individuals receive information, it will exceed their social networks national identity similarity since the value of *s* is 0.8, leaving 0.3 difference which means that only those neighbours whose national identity similarity overlaps would be able to exert social influence. Hence why we observe such strong patterns of bi-polarisation at the highest values of *s*. As individuals interact with each other and exchange information, those at the extreme will draw in those initially at the centre because of the homophily principle and the majority and minority national identity biases (see Karimi et al. 2022; Neuhauser et al. 2022).

The findings from the sensitivity analyses suggest that indeed, national identity dynamics are sensitive to changes in the bounded confidence threshold of national identity similarity, *s*, like previous OD models have shown (see Deffuant et al. 2008; Amblard and Deffuant 2004; Mäs et al. 2010). Depending on the initial social network and social media filter bubble scenarios, certain values of *s* produced different national identity dynamics from convergence to bi-polarisation. More specifically, the presence of social media filter bubbles promoted polarisation of national identity national identities the larger the value of *s* suggesting that getting exposed to information with a high national identity-similarity threshold would actually promote bi-polarisation. Furthermore, the initial social network condition had a significant impact on the resulting national identity dynamics, at the highest values of the bounded confidence threshold of national identity similarity, *s*. Homophilous social networks, representing social networks where the majority of one's social connections shared a similar national identity, promoted national identity polarisation, in combination with filter bubbles, at lower values of *s* compared to initially random networks. In other words, online echo chambers, found primarily on homophilous social networks, speed up the national identity polarisation process promoted by social media filter bubble's selective exposure.

Acknowledgements

I would like to thank Dr. Stefano Picascia for his insightful theoretical discussions, computational support, and comments on this paper; Prof. Sergi Pardos-Prado for his theoretical support and feedback provided on this paper; Dr. Eric Silverman for the early discussions during my PhD; the agent-based modelling group and the systems science teams at the Social and Public Health Sciences Unit at Glasgow University for their helpful discussions; and the Comparative Politics Cluster at Glasgow University for their time and feedback on previous versions of this manuscript. Lastly, I would also like to thank the anonymous reviewers of the Social Simulation Conference 2022 and Dr. Jennifer Badham for their comments that greatly improved this manuscript.

Author contributions

CCC designed and implemented the agent-based model simulation. CCC analysed and interpreted the simulation results. This manuscript was written in its entirety by CCC.

Funding

This research was funded by the Economic and Social Research Council (ES/P000681/1) and was carried out at University of Glasgow as part of the main author's (CCC) PhD.

Availability of data and materials

The source code, primary and secondary data used in the model, and documentation can be found on Github under a GNU General Public License 3.0. All data produced by the simulations can be accessed in OSF Repository under a Creative Commons License By Attribution 4.0 International.

Declarations

Competing interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Received: 22 February 2024 Accepted: 14 October 2024
Published online: 11 November 2024

References

Albert R, Barabási A (2002) Statistical mechanics of complex networks. Rev Modern Phys 74(1):47–97
Almagor J, Picascia S (2020) Exploring the effectiveness of a COVID-19 contact tracing app using an agent-based model. Sci Rep 10:1–11
Amblard F, Deffuant G (2004) The role of network topology on extremism propagation with the relative agreement opinion dynamics. Phys A: Stat Mech Appl 343:725–738
Arnaboldi V, Conti M, La Gala M et al (2016) Ego network structure in online social networks and its impact on information diffusion. Comput Commun 76:26–41
Axelrod R (1997) The Complexity of Cooperation, 1st edn. Princeton University Press, Princeton
Badham J (2013) Commentary: Measuring the shape of degree distributions. Netw Sci 1(2):213–225
Bail C, Argyle LPA, Brown TW, and Volfovsky A (2018) Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci (PNAS) USA 115:9216–9221
Bakshy E, Messing LS, Adamic (2015) Exposure to ideologically diverse news and opinion on facebook. Science 348:1130–1132
Baldassarri D, Bearman P (2007) Dynamics of political polarization. Am Sociol Rev 72:784–811
Balietti S, Getoor L, Goldstein D et al (2021) Reducing opinion polarization: effects of exposure to similar people with differing political views. Proc Natl Acad Sci PNAS) 118(52):1–11
Bandy J, Diakopoulos N (2021) Curating quality? How Twitter's timeline algorithm treats different types of news. Soc Media Soc 7:1–17
Banisch S, Olbrich E (2019) Opinion polarization by learning from social feedback. J Math Sociol 43:76–103
Barabási AL (2009) Scale-free networks: A decade and beyond. Science 325:412–13
Barberá P, Jost J, Nagler J et al (2015) Tweeting from left to right: Is online political communication more than an echo chamber? Psychol Sci 26:1531–1542
Barberá P (2014) Birds of the same feather tweet together. bayesian ideal point estimation using twitter data. Polit Anal pp 1–56
Baumann F, Lorenz-Spreen P, Sokolov IM et al (2020) Modeling echo chambers and polarization dynamics in social networks. Phys Rev Lett 124:048301
Beam MA, Hutchens MJ, Hmielowski JD (2018) Facebook news and (de)polarization: Reinforcing spirals in the 2016 US election. Inf Commun Soc 21:940–958
Boccaletti S, Bianconi G, Criado R et al (2014) The structure and dynamics of multilayer networks. Phys Rep 544(1):1–122
Bos L, Kruikemeier S, de Vreese C (2016) Nation binding: How public service broadcasting mitigates political selective exposure. PLOS ONE 11:1–15
Boutyline A, Willer R (2016) The social structure of political echo chambers: Variation in ideological homophily in article networks. Polit Psychol 38:1–52
Bozdag E (2013) Bias in algorithmic filtering and personalization. Ethics Inf Technol 15:209–227
Centre d'Estudis d'Opinion (CEO) (2022) Baròmetre d'Opinió Política (BOP), 1a onada 2011, Generalitat De Catalunya (ed), In: Baròmetre d'Opinió Política Series, Dossier de Prensa, accessed 21/04/2022, https://bit.ly/3DCnvcl
Cardenal A, Aguilar-Paredes C, Galais C et al (2019) Digital technologies and selective exposure: how choice and filter bubbles shape news media exposure. Int J Press/Polit 24:465–486
Chaudhry I, Gruzd A (2020) Expressing and challenging racist discourse on facebook: How social media weaken the "spiral of silence" theory. Policy internet 12:88–108
Chen W, Pacheco D, Yang K et al (2021) Neutral bots probe political bias on social media. Nat Commun 12:1–10

Cioroianu I, Banducci S, Jasny L, et al (2018) To polarize or not: comparing networks of news consumption. In: CeDEM Asia 2018: Proceedings of the International Conference for E-Democracy and Open Government. Edition Donau-Universität Krems, p 77–97
Dahlgren PM (2019) Selective exposure to public service news over thirty years: the role of ideological leaning, party support, and political interest. Int J Press/Polit 24:293–314
Deffuant G (2006) Comparing extremism propagation patterns in continuous opinion models. JASSS, 9:1–10
Deffuant G, Huet S, Amblard F (2005) An individual-based model of innovation diffusion mixing social value and individual benefit. Am J Sociol 110:1041–1069
Deffuant G, Neau D, Amblard F et al (2000) Mixing beliefs among interacting agents. Adv Complex Syst 3:87–98
Deffuant G, Amblard F, Weisbuch G et al (2002) How can extremism prevail? A study based on the relative agreement interaction model. JASSS 5(4)
Deffuant G, Skerrat S, Huet S (2008) An agent based model of agri-environmental measure diffusion: What for? In: Lopez-Paredes A, Hernandez Iglesias CE (eds) Agent based modelling in natural resource management, p 55–73
Del Vicario M, Bessi A, Zollo F et al (2016) The spreading of misinformation online. Proc Natl Acad Sci PNAS) 113:554–559
Del Vicario M, Scala A, Caldarelli G et al (2017) Modeling confirmation bias and polarization. Sci Rep 7,4031
Dubois E, Blank G (2018) The echo chamber is overstated: the moderating effect of political interest and diverse media. Inf Commun Soc 21(5):729–745
Dunbar R, Arnaboldi V, Conti M et al (2015) The structure of online social networks mirrors those in the offline world. Soc Netw 43:39–47
Elliott J (2018) Scots and Catalans, union and disunion. Yale University Press, London
Elsenbroich C, Anzola D, Gilbert N (2016) Social dimensions of organised crime: modelling the dynamics of extortion rackets. Springer
Epstein J (2002) Modeling civil violence: An agent-based computational approach. Proc Natl Acad Sci PNAS) 99:7243–7250
Epstein J (2013) Agent_Zero: Toward neurocognitive foundations for generative social science. Princeton University Press, Princeton
Festinger L (1957) A theory of cognitive dissonance. Row Peterson, Evanston
Finkel CEJ, Bail Cikara M, Ditto P et al (2020) Political sectarianism in America. Science 370(6516):533–536
Flache A (2018) Between monoculture and cultural polarization: Agent-based models of the interplay of social influence and cultural diversity. J Archaeol Methods Theory 25:996–1023
Flache A, Macy M (2011) Small worlds and cultural polarization. J Math Sociol 35:146–176
Flache A, Mäs M (2008) How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams. Comput Math Org Theory 14:23–51
Flache A, Mäs M, Feliciani T et al (2017) Models of social influence: towards the next frontiers. JASSS, 20:1–31
Flaxman S, Goel S, Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. Public Opin Qly 80:298–320
Fletcher R, Robertson C, Nielsen RK (2021) How many people live in politically partisan online news echo chambers in different countries?. J Quant Descr Digit Media 1:1–56
Fletcher R, Kalogeropoulos A, Nielsen R (2021a) More diverse, more politically varied: How social media, search engines and aggregators shape news repertoires in the united kingdom. New Media & Society pp 1–22
Fowler J, Smirnov O (2005) Dynamic parties and social turnout: An agent-based model. Am J Sociol 110(4):1070–1094
Fränken J, Pilditch T (2018) Cascades across networks are sufficient for the formation of echo chambers: an agent-based model. J Artif Soc Soci Simul JASS) 24:1–5
Garrett RK, Carnahan D, Lynch EK (2013) A turn toward avoidance? Selective exposure to online political information, 2004–2008. Polit Behav 35:113–134
Garrett RK, Stroud NJ (2014) Partisan paths to exposure diversity: Differences in pro- and counterattitudinal news consumption. J Commun 64:680–701
Geschke D, Lorenz J, Holtz P (2019) The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. Br J Soc Psychol 58:129–149
Gibson J, Sutherland J (2020) Keeping your mouth shut: Spiraling self-censorship in the united states. PNAS pp 1–20
Gil de Zúñiga H, Diehl AT, Ardévol-Abreu (2017) Internal, external, and government political efficacy: Effects on news use, discussion, and political participation. J Broadcast Electron Media 61:574–596
Gilbert N (2008) Agent-Based Models. Sage Publications, London
Gil de Zuñiga H, Cheng Z, Gonzalez-Gonzalez P (2022) Effects of the news finds me perception on algorithmic news attitudes and social media political homophily. J Commun 0:1–14
Goel S, Mason W, Watts D (2010) Real and perceived attitude agreement in social networks. J Personal Soc Psychol 99:611–621
Granovetter M (1973) The strength of weak ties. Am J Sociol 78:1360–1380
Granovetter M (1978) Threshold models of collective behavior. Am J Sociol 83:1420–1443
Guess A (2021) (almost) everything in moderation: new evidence on americans' online media diets. Am J Polit Sci 65:1007–1022
Guess A, Nagler J, Tucker J (2019) Less than you think: Prevalence and predictors of fake news dissemination on facebook. Sci Adv 5:1–5
Guess A, Nagler J, Tucker J (2020) Less than you think: Prevalence and predictors of fake news dissemination on facebook. J Polit 82:1423–1440
Gómez S, Díaz-Guilera A, Gómez-Gardeñes J et al (2013) Diffusion dynamics on multiplex networks. Phys Rev Lett 110:1–6
Halberstam Y, Knight B (2014) Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. J Public Econ pp 1–40
Hegselmann R, Krause U (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. JASSS, 5:1–33

Hegselmann R, Krause U (2004) Opinion dynamics driven by various ways of averaging. Comput Econ 25:381–405
Holtug N (2020) National identity, shared values, and social cohesion. In: Me Sardoč (ed) Handbook of patriotism. Springer, Cham, pp 499–511
Iyengar S, Hahn K (2009) Red media, blue media: Evidence of ideological selectivity in media use. J Commun 59:19–39
Iyengar S, Sood G, Lelkes Y (2012) Affect, not ideology: a social identity perspective on polarization. Public Opinion Qly 76:405–431
Jager W, Amblard F (2005) Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. Comput Math Org Theory 10:295–303
Karimi F, Oliveria M, Strohmaier M (2022) Minorities in networks and algorithms. Phys Soc pp 1–16
Keijzer M (2022) Opinion dynamics in online social media. PhD thesis, Groningen University, available at https://research.rug.nl/en/publications/opinion-dynamics-in-online-social-media
Keijzer M, Mäs M (2022) The complex link between filter bubbles and opinion polarization. Data Sci pp 1–28
Kemp S (2022) Digital 2021: Spain. DataReportal on behalf of Kelpios Accessed 28/02/2022, https://datareportal.com/reports/digital-2021-spain
Kim M, Lu Y (2020) Testing partisan selective exposure in a multidimensional choice context: Evidence from a conjoint experiment. Mass Commun Soc 23:107–127
Kinsley A, Rossi G, Silk M et al (2020) Multilayer and multiplex networks: an introduction to their use in veterinary epidemiology. Front Vet Sci 7:1–13
Kumpel A, Karnowski V, Keyling T (2015) News sharing in social media: A review of current research on news sharing users, content, and networks. Social Media Soc 1:1–14
Latané B (1996) Dynamic social impact: the creation of culture by communication. J Commun 46:13–25
Lazarsfeld P, Berelson B, Gaudet H (1948) The people's choice, 2nd edn. Columbia University Press, New York
Lelkes Y (2016) Mass polarization: Manifestations and measurements. Public Opinion Qly 80:392–410
Levendusky MS (2013) Why do partisan media polarize viewers?. Am J Polit Sci 57(3):611–623
Li J, Xiao R (2017) Agent-based modelling approach for multidimensional opinion polarization in collective behaviour. JASSS, 20(2):4
Li Y, Zhao H (2021) Evaluation and control of opinion polarization and disagreement: a review. Int J Crowd Sci 5:1–15
Lindez-Borrás M (2013) Movilización en la sociedad catalana: aparición y pervivencia. Clivatge Revista Cientifica (in Spanish) 2:1–29
Lodge M, Taber C (2000) Three steps towards a theory of motivated reasoning. In: Lupia A, McCubbins M, Popkin S (eds) Elements of reason: cognition, choice and the bounds of rationality. Cambridge University Press, New York
Lubbers M, Molina J, Valenzuela-García H (2019) When networks speak volumes: Variation in the size of broader acquaintanceship networks. Soc Netw 56:55–69
Macy MW, Kitts JA, Flache A, et al (2003) Polarization in dynamic networks: A hopfield model of emergent structure. In: R. Breiger R, Carleyand K, Pattison PE (eds) Dynamic social network modeling and analysis: workshop summary and papers, pp 162–173
Maes M, Bischofberger L (2015) Will the personalization of article social networks foster opinion polarization? Social Science Research Network (SSRN) pp 2–36
Masip P, Suau J, Ruiz-Caballero C (2020) Incidental exposure to non-like- minded news through social media: opposing voices in echo-chambers' news feeds. Media Commun 8:53–62
McClain C (2021) 70% of U.S. social media users never or rarely post or share about political, social issues. Pew Research Center Publications Accessed on 16/01/2024 from https://www.pewresearch.org/short-reads/2021/05/04/70-of-u-s-social-media-users-never-or-rarely-post-or-share-about-political-social-issues/
McPherson M, Smith-Lovin L, Cook J (2001) Birds of a feather: Homophily in social networks. Annu Rev Sociol 27:415–444
Metzger M, Flanagin A (2013) Credibility and trust of information in online environments: the use of cognitive heuristics.J Pragmat 59:210–220
Metzger M, Flanagin A, Medders R (2010) Social and heuristic approaches to credibility evaluation online. J Commun 60:413–439
Milli L (2021) Opinion dynamic modeling of news perception. Appl Netw Sci 6:1–19
Moreno L (1995) Multiple ethnoterritorial concurrence in Spain. National Ethnic Polit 1:11–32
Mäs M, Flache A (2013) Differentiation without distancing, explaining bi-polarization of opinions without negative influence. PLoS ONE 8:1–17
Mäs M, Flache A, Helbing D (2010) Individualization as driving force of clustering phenomena in humans. PLoS Comput Biol 6
Neuhauser L, Karimi F, Bachmann J, et al (2022) Improving the visibility of minorities through network growth interventions. Physics and Society (physicssoc-ph) ArXiv:2208.03263v1:1–18
Noelle-Neumann E (1974) The spiral of silence a theory of public opinion. J Commun 24:43–51
Nyhan B, Settle J, Thorson E et al (2023) Like-minded sources on facebook are prevalent but not polarizing. Nature 620:137–144
Osmundsen M, Bor A, Ahlstrup P et al (2021) Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. Am Polit Sci Rev 115:999–1015
Ovadya A, Thorburn L (2023) Bridging systems: Open problems for countering destructive divisiveness across ranking, recommenders, and governance. ArXiv pp 1–39
Pardos-Prado S, Dinas E (2010) Systemic polarisation and spatial voting. Eur J Polit Res 49:759–786
Parella J (2015) La secesión como solución al problema del estado, el caso de Cataluña. Procesos de Mercado: Revista Europea de Economía Política (in Spanish) 12:237–275
Pariser E (2011) The filter bubble: what the internet is hiding from you. Penguin Group, London
Pennycook G, Rand D (2019) Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition 188:39–50
Rafaeli S, Ravid G, Soroka V (2004) De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. In: 34th Hawaii international conference on system sciences. Hawaii

Rau P, Gao Q, Ding Y (2008) Relationship between the level of intimacy and lurking in online social network services. Comput Hum Behav 24:2757–2770
Richards B (2013) National identity and social cohesion: theory and evidence for British social policy. PhD thesis, The London School of Economics, available at https://etheses.lse.ac.uk/910/1/Richards_National-identity-and-social-cohesion_2013.pdf
Romero-Moreno G, Padilla J, Brede M (2021) The effects of party competition on consensus formation. In: Conference: 3rd international workshop on agent-based modelling of human behaviour (ABMHuB'21)
Ross-Arguedas A, Robertson C, Fletcher R, et al (2022) Echo chambers, filter bubbles, and polarisation: a literature review. Reuters Institute for the Study of Journalism Research Pieces
Sayama H (2020) Complex systems in a nutshell. In: Sayama H (ed) Introduction to the modeling and analysis of complex systems. Binghamton University, State University of New York
Schweighofer S, Garcia D, Schweitzer F (2020) An agent-based model of multi-dimensional opinion dynamics and opinion alignment. Chaos Interdiscipl J Nonlinear Sci 30(9)
Seawright J, Gerring J (2008) Case selection techniques in case study research: a menu of qualitative and quantitative options. Polit Res Qly 61:294–308
Settle J (2018) Frenemies: how social media polarizes America. Cambridge University Press
Sibona C, Walczak S (2011) Unfriending on facebook: friend request and online/offline behavior analysis. In: 44th Hawaii International Conference on System Sciences 2011. Institute of Electrical and Electronics Engineers (IEEE), p 1–10
Singh V, Basak S, Modanwal N (2011) Agent based modeling of individual voting preferences with social influence. In: Nagamalai D, Renault E, Dhanuskodi M (eds) Trends in computer science, engineering and information technology. Springer, Berlin Heidelberg, pp 542–552
Smith M, Rainie L, Shneiderman E, et al (2014) Mapping twitter topic networks: From polarized crowds to community clusters. Pew Research Center
Stroud NJ (2008) Media use and political predispositions: Revisiting the concept of selective exposure. Polit Behav 30:341–366
Taber CS, Lodge M (2006) Motivated skepticism in the evaluation of political beliefs. Am J Polit Sci 50:755–769
Taber CS (2003) Information processing and public opinion. In: Sears D, Jervis R, Huddy L (eds) The Oxford handbook of political psychology. Oxford University Press, London
Takács K, Mäs M, Flache A (2016) Discrepancy and disliking do not induce negative opinion shifts. PLoS ONE 11:1–21
Trilling D, Schoenbach K (2013) Skipping current affairs: the non-users of article and offline news. Eur J Commun 28:35–51
Tucker J, Guess A, Barberá P, et al (2018) Social media, political polarization, and political disinformation: a review of the scientific literature. Social Science Research Network (SSRN) pp 1–95
Vaccari C, Valeriani A, Barberá P et al (2016) Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of twitter. Social Media + Soc 2:1–16
Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. Nature 393:440–442
Weeks B, Holbert R (2013) Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship. J Mass Commun Qly 90:212–232
Wilensky U, Rand J (2015) An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo. The MIT Press, Cambridge
Wilensky U (1999) NetLogo (and NetLogo User Manual), http://ccl.northwestern.edu/netlogo/ edn. Center for Connected Learning and Computer-Based Modeling: Northwestern University
Williams H, McMurray J, Kurz T et al (2015) Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global Environm Change 32:126–138
Zuiderveen Borgesius F, Trilling D, Möller J et al (2016) Should we worry about filter bubbles?. Internet Policy Rev 5:1–15

Publisher's Note