The triple filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers

Daniel Geschke[1], Jan Lorenz[2], & Peter Holtz[3]

[1] Institut für Demokratie und Zivilgesellschaft (Institute for Democracy and Civil Society, IDZ) Jena, Germany

[2] Jacobs University Bremen, BIGSSS Bremen International Graduate School of Social Sciences, Germany

[3] Leibniz-Institut für Wissensmedien IWM (Knowledge Media Research Center), Tübingen, Germany

Submitted at the British Journal of Social Psychology (BJSP); December 29 2017.

Running Head: The Triple Filter Bubble

Contact information of the corresponding author:

Dr. Peter Holtz, IWM Leibniz-Institut für Wissensmedien, Schleichstraße 6, 72076 Tübingen, Germany (e-mail: p.holtz@iwm-tuebingen.de; phone: +49 7071 979-303).

Requests for reprints should be addressed to Daniel Geschke daniel.geschke@idz-jena.de

Abstract

A theoretical framework for the emergence of filter bubbles and echo chambers is tested using agent-based modelling. It is hypothesized that information filtering processes take place on the individual, the social, and the technological levels (triple-filter-bubble framework). An agent-based model with nine different information filtering scenarios was constructed to study their respective effects on individuals and society and further, to answer the question under which circumstances the Web 2.0 with its social and technological filtering processes leads to fragmentation of modern society into distinct echo chambers or opinion groups. Simulations show that, even without any social or technological filters, echo chambers emerge as a consequence of cognitive mechanisms, such as confirmation bias, under conditions of central information propagation through channels reaching a large part of the population. When social and technological filtering mechanisms are added to the model, polarisation of society into even more distinct and less inter-connected echo chambers is observed. Merits and limits of the theoretical framework, and more generally of studying complex social phenomena using agent-based modelling, are discussed. Directions for future research and possible measures against societal fragmentation are suggested.

The triple filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers

The ubiquitous availability of information in the age of social media and the personalization of information flows have had substantial effects on our daily lives and on our socio-political culture (Castells, 2010; Hermida, Fletcher, Korell, & Logan, 2012; Happer & Philo, 2013). Still, there is disagreement if and under which circumstances these technological changes have positive or negative consequences for individual users and society as a whole. Whereas some optimists are confident that the internet and the social media expand everyone's chances to find unbiased information (e.g., Michal Kosinski in Noor, 2017), critics have warned of the emergence of *filter bubbles*, minimizing exposure to information which challenges individual attitudes (Pariser, 2011). Filter bubbles are defined here as an individual outcome of different processes of information search, perception, selection, and remembering, the sum of which causes the perceived information to fit one's pre-existing attitudes. On the societal level, individuals *sharing* a common social media bubble comprised of internet content which confirms certain ideologies are prone to processes of group radicalization and polarization (Vinokur & Burnstein, 1978); this phenomenon has come to be known as *echo-chamber effect* (Sunstein, 2001 & 2009; Garrett, 2009). Thus, echo chambers are a social phenomenon where the filter bubbles of interacting individuals strongly overlap.

In our study, we analyse the interplay between processes on the levels of individual minds, social groups, and technology, using agent-based modelling (Smith & Conrey, 2007). Empirical studies of the effects of social media or other technological developments on individual human beings, social groups, and societies always suffer from the problem that social media effects develop and accumulate over a substantial time-span and involve social activities such as chatting and content sharing. For these reasons, it is difficult to model such effects in psychological laboratory experiments. Furthermore, technical revolutions, such as the emergence of Web 2.0, affect more or less every member of a given society within a relatively short time-span. It is therefore also difficult to use longitudinal or quasi-experimental designs to study the effects of technology on individual users (Livingstone, 1996). The goal of our study is first to summarize relevant existing theories on filter bubbles and echo chamber effects in a formal agent based model and then to explore the dynamics of this model. This approach enables us to distinguish individual effects from effects which emerge in their interaction with macroscopic phenomena (Flache, Mäs, Feliciani, Chattoe-Brown, Deffuant, Huet, & Lorenz, 2017) and in turn, to tentatively gauge possible effects of changing technological environments on individuals and society.

In our study, we employ an interdisciplinary approach, in so far as our models take into account theories from cognitive psychology, social psychology and micro-sociology. The technique of agent-based models permits computer-based model building and simulations of social processes of interacting individuals, as a means of studying links between the micro level of individual behaviour and any macro level effects which emerge. The method has been used to simulate such different social phenomena as the occurrence of traffic jams (Bazzan & Klügl, 2014) or segregation in housing (Huang, Parker, Filatova, & Sun, 2014). In the next paragraph, we will briefly summarize the available empirical research on filter bubble and echo chamber effects and afterwards present our triple filter bubble framework.

Empirical findings on filter bubbles and echo chambers

Already in 2009, Garrett had found evidence that within the political domain, internet users preferred to consume information which confirmed their ideologies. Del Vicario and colleagues (Del Vicario et al., 2016) were recently able to show in a study using Facebook data how information is passed along ideological fault lines in scientific as well as conspiracy-theory communities. Bakshy, Messing, and Adamic (2015) analysed the data of 10.1 million US-American Facebook users who identified themselves as being either politically liberal, moderate, or conservative. They found that most information filtering is the result of *homophily,* in the sense that Facebook users have significantly more friends with a political orientation similar to their own. The Facebook news feed then relies on information that was shared by at least one person in the friend network, and this already leads to information selection with a severe bias in favour of information confirming a certain ideology. However, an earlier study of the same research group (Bakshy, Rosenn, Messing, & Adamic, 2012) found that the majority of the information that was displayed in the Facebook newsfeed was not shared by close friends with whom the Facebook users exchanged chats and comments and updates on a regular basis (*strong ties*), but by acquaintances with whom users only communicated on a casual basis (*weak ties;* cf. Granovetter, 1973). As long as there is at least some heterogeneity within a user's friend network, the user will at least have some exposure to differing points of view. Such information would be totally out of the user's reach if information were only accessible via offline communication with close friends who normally share a person's ideologies and beliefs.

A somewhat related result was found using Twitter data (Vaccari, Valeriani, Barbera, Jost, Nagler, & Tucker, 2016): On the one hand, Twitter users more frequently interact (comment or retweet) with authors with a similar political ideology; still, Twitter is used frequently as well to interact with representatives of networks that display an oppositional ideology. The authors draw the conclusion that apparently there are not only echo chambers on Twitter, but also *contrarian clubs*. However, another group of researchers found a definite longitudinal political polarization of Twitter users: Over time (between 2009 and 2016), the number of politicians and media with similar ideologies that the users followed increased continuously (Garimella & Weber, 2017).

Recently, the specific role of recommender systems in the emergence of filter bubbles and echo chambers has been investigated in a number of empirical studies. Technically speaking, recommender systems provide recommendations in three basic ways: They recommend content that was previously selected by a user or other (to some extent similar) users (*collaborative filtering*). They recommend content based on similarities of properties and characteristics between previously chosen and available content (*content-based filtering*). Or they combine both approaches (*hybrid recommender systems*; Burke, 2002). Nguyen et al. (Nguyen, Hui, Harper, Terveen, & Konstan, 2014) found mixed results when they analysed the effects of a movie rating page's collaborative filtering based recommender system on its users' range of interests. The users' average movie diversity decreased over time, but the effect was stronger for those users that did not usually follow the recommendations than for those who did frequently click the recommended links. In a study on the effects of a music platform's recommender system, Hosanagar and colleagues found little empirical evidence for fragmentation over time (Hosanagar, Fleder, Lee, & Buja, 2013). The development of recommendation strategies that counter possible filter bubble or echo chamber effects has become a topic of interest for software engineers in the last several years (e.g., Resnick, Garrett, Kriplean, Munson, & Stroud, 2013).

Overall, there seems to be no common interpretation of the available evidence in the research community as to whether technological features such as recommender systems or many-to-many communication patterns in social media facilitate or attenuate the emergence of filter bubbles and echo chambers.

The triple-filter-bubble model

We refer to *filters* in a very general way as processes that lead to a limitation of information that is available to individuals. In our models, we will take into account filtering processes on three different levels: The individual, the social, and the technological level (Geschke, 2017).

Individual filters: Cognitive-motivational processes. The first group of filters – cognitive motivational processes –has been studied extensively in cognitive and social psychology. As a means of confirming pre-existing attitudes (Nickerson, 1998), avoiding cognitive dissonance (Festinger, 1957), and boosting social identity (Brewer, 1991), individuals are to different extents cognitively motivated to search for and add fitting bits of information and to ignore or deny conflicting ones. Here, filtering refers to selective exposure due to an individual's information search, processing, and memory. Curiosity may, however, motivate individuals to have a preference for consuming information that is at least to some degree novel and surprising (Loewenstein, 1994).

Social filters: Group dynamic communication processes. The second group of filters – *group dynamic communication processes* – works on the social level. The *group polarization* effect (Vinokur & Burnstein, 1978), for example, leads to increasingly similar attitudes of individual group members after discussions. These attitudes tend to be more extreme than the average individual attitudes before the group discussion. We expect such effects to emerge from social media interactions among like-minded individuals as well.

Another type of group effects to consider is related to human beings' need for *conformity* (Asch, 1951; Hogg & Turner, 1987). Here, users wish to comply with and to internalize information that corresponds to the (perceived) majority view, which is supposedly usually moderate rather than extreme. Thus, in the case of widely shared and discussed information, majority influence could lead to reduced polarization of attitudes.

Technological filters: Algorithms. The third group of filters – *algorithms* – operates on the technological level: Online media providers, such as Google or Facebook, compete for user attention. Therefore, they filter the provided information according to individual users' assumed wants and needs, leading to individually selected media offers (Pariser, 2011). The goal of this filtering is to maximise the time users spend on their respective sites, in order to maximise profits generated through advertising. To accomplish this, these private companies use proprietary, non-transparent automatic algorithms. In effect, this leads to different information offers tailored to the individual. For instance, none of us gets the same output on any given Google search; instead, each user gets an individualised selection of information. We assume that stronger automatic filtering leads to a decreased variety of information that is offered to individuals. This eventually leads to a decreased spectrum of attitudes that are cognitively available and salient in individuals, and thus, to smaller filter bubbles.

However, these recommender systems also constantly confront the user with previously unknown novel information to maximize click-through rates, thereby potentially increasing the exposure to different points of view (Herlocker, Konstan, Terveen, & Riedl, 2004). Therefore, an alternative assumption is that, in spite of the filtering processes mentioned above, online media increase the spectrum of attitudes that are cognitively available and salient in individuals.

In sum, the filters on these three levels are expected to influence how much of the abundant information is cognitively available to individuals. More importantly, this influence is not random, but systematic: Information is more likely sought, delivered or perceived when it fits the individual's needs as determined by individual characteristics, and this is partly gauged through automatic recommender systems. Additionally, the outcome depends as well on attitudinal characteristics of the peer groups that individuals interact with.

Research Questions

We have modelled filtering processes to find out how traditional, central media channels (one-to-many communication) and modern social networks of Web 2.0 (many-to-many communication), as well as automatic recommender systems influence attitude formation in their different ways. The effects of nine different scenarios of information propagation are compared. Each scenario represents the influence of a different combination of filtering mechanisms. Their effects are gauged on the formation of echo chambers, as represented by closely knit communities (clusters of people in the information space) that share the same information and integrate little new outside information. We created a parsimonious, agent-based model to simulate these different processes.

Methods

Agent-based modelling

In agent-based models, interactions between individual agents are simulated as a consequence of rules that were set by the researcher, and effects which emerge on a macro level are explored as outcomes. Therefore, this method is perfectly suited to test the current research question of how individual information search, posting, and remembering, as well as group processes and technological filtering together affect societal polarisation and fragmentation. Agent-based models are particularly suited for studying interactions between individuals instead of interactions between variables (Smith & Conrey, 2007) and how such interpersonal influence processes play out (Mason, Conrey, & Smith, 2007).

Model synopsis

We designed a dynamic agent-based model where several individuals (together representing a society) position themselves in a two-dimensional attitude space based on info-bits they hold in memory. Individuals repeatedly receive new information with differing attitudinal messages from different sources. The sources of new information can be a) central announcement, e.g. through mass media, b) individual discovery, or c) personalized recommendations, e.g. through online media providers. When social media mechanisms are provided, individuals may also d) post information to friends in their social network. Individuals integrate information through cognitive processes: They integrate a particular info-bit more likely when the distance of its attitudinal message to their own attitude is below the threshold of their latitude of acceptance. This means that it is unlikely that they integrate information that does not fit their pre-existing average attitudes. Individuals have a limited memory for a certain number of info-bits; when their memory is full and they integrate a new info-bit, they have to forget a random one. These processes lead to repositioning of individuals in the attitudinal space according to the average information they consequently hold in their memory.

The model includes social networks with different predefined numbers of communities in which individuals are tightly linked, while they have few links to members of other communities. Further on, the thresholds for the maximal number of info-bits in memory and the latitude of acceptance can be *blurred* in such a way that forgetting also sometimes happens below the thresholds or may even occasionally not happen at all beyond the thresholds. Finally, the turnover of individuals through new individuals' dropping-out and entering can be switched on in the model. Social network dynamics can take place, where individuals can form new friendships with a friend of a friend for the sake of unfriending a friend when the attitudinal distance is too great.

Entities and Initial Conditions

The entities in the model are:

- *Individuals:* We consider a fixed number, even when there is some turnover as individuals drop out and new ones enter. Individuals are characterized by their attitudinal position in the two-dimensional attitude space which can change over time, and by the label of their friendship community which stays constant throughout its lifetime.
- *Friend-links* are undirected links between individuals which represent the social network between them. The number of friend-links is held constant after initialization even when social network dynamics are switched on.
- *Info-bits* appear and vanish as points in the same attitudinal space where individuals position themselves. They represent news items from mass media, recommended from online media providers, or produced by individuals, stripped down to their attitudinal message.
- *Info-links* are undirected links between individuals and info-bits, which represent the fact that the individual holds a particular info-bit in memory. Info-links create a bipartite network between individuals and info-bits. Thus, they also create indirect connections between individuals who link to (i.e. know) the same info-bit. Such a link between individuals will be called *info-sharer link*.

In the following section we define all parameters and variables (in `courier typeface`) of the model. We give baseline numbers and conditions for all parameters including initial conditions. This creates the baseline setting for the simulation experiments.

- The *attitude space* is a square in two-dimensional Euclidean space with the origin at its centre. We chose a parametrization where attitudes can have values between -16 and 16 in both attitude dimensions. Initially, `numguys` individuals (baseline 500) are created with random positions in the attitude space (uniformly distributed).
- A *social network* of friend-links is created between individuals characterized by an expected number of friends `numfriends` (baseline 20). We construct random networks with the community structure quantified by the number of communities `numcommunities` (baseline 4). For the case `numcommunities`=1 this equals a classical random graph where a link between any two individuals exists with the probability `numfriends/numguys` (baseline 20/500=0.04). When the expected number of friends is larger than one, all or almost all individuals are indirectly connected to all other individuals through friendship links. Although this can also be found in real-world social networks, real networks usually show much more clustering or community structure than a random graph (Newman, 2003). Therefore, we also allow for random social networks with `numcommunities` communities. The assignment of community labels to individuals is random with uniform probability. This results in communities of equal size in expectation. The size of a community follows a binomial distribution with success probability 1/`numcommunities` and `numguys` number of draws (baseline ¼ and 500, leads to M=125 and SD=9.68). A link between two individuals is created such that the expected total number of friends of each is `numfriends` and a certain fraction of these are links to individuals of the

same community (`fraction-intra` with baseline 0.8). This set-up is analogous to a random network. For any two nodes a link is created either with probability `p-intra` when the two individuals are from the same community or with probability `p-inter` when they are from different communities. The probabilities `p-intra` and `p-inter` can be computed from the parameters `numguys`, `numfriends`, `numcommunities`, and `fraction-intra`. The number of friends follows a binomial distribution. This social network is independent of initial attitudes and only subject to change when the probability of refriending is larger than zero.

Integration of new information

The core cognitive process of individuals is the integration of new information. Over time individuals perceive new bits of information. They integrate such an info-bit into memory based on its attitudinal message characterized by its position in the attitude space. Integration is a binary random event based on the `integration-probability`, which is a function of the attitude distance between the individual and the info-bit d. An info-bit is integrated with certainty if the attitudinal position of the info-bit coincides with the attitudinal position of the individual (d=0). The probability of integrating an info-bit decreases with d (cf. Fisher & Lubin 1958, Abelson 1964, Fishbein & Ajzen 1975). This means that information fitting the individual's average pre-existing attitudes is more likely to be integrated. The decrease is shaped by two parameters, the *latitude of acceptance D (*Sherif & Hovland, 1961*)*, and the sharpness parameter δ which specifies how sharply the integration probability drops from one to zero around the latitude of acceptance. We use the following functional formula

$$P(d; D, \delta) = \frac{D^\delta}{d^\delta + D^\delta} \qquad (1)$$

This formula follows the formalization of the *Social Judgment Theory* of Hunter, Danes, and Cohen (1984). However, they only dealt with the case of δ=2 and did not take into consideration a sharper decline around the latitude of acceptance. The limit in the case of very large δ coincides with the bounded confidence model (Deffuant, Neau, Amblard, & Weisbuch, 2000; Krause, 2000). Latitude of acceptance *D* determines at which distance the integration probability is 0.5. At the limit of very large sharpness parameters, probabilistic remembering becomes deterministic. In this case, info-bits are integrated with certainty if the distance is less than *D*, and are rejected otherwise. As a baseline we use a latitude of acceptance *D*=5 and sharpness which is close to deterministic δ=20. The function of the parameters *D* and δ are shown in Figure 1 below.

< Figure 1 >

Individuals can only have info-links to a certain maximum number of `memory` info-bits. As a baseline case we use `memory`=20. When an individual has a full memory and wants to integrate a new info-bit, a random info-link is dropped (i.e., *forgotten*) before the new info-bit is integrated. After the integration of a new info-bit, the individual readjusts its attitudinal position to the average attitude of the info-bits she holds in memory, which follows Anderson's (1971) *integration theory*.

Model dynamics

The dynamics of the model evolve in discrete events grouped in time steps called ticks. In each tick the following events take place:

1. **New info-bits.** There are four modes of creating new info-bits. In the *central* mode, one info-bit is created at random in the attitude space and every individual tries to integrate this info-bit. This represents mass media input from one central, unbiased channel

(one-to-many communication). In the *individual* mode, each individual creates one info-bit at a random position and tries to integrate it. This mode represents the process in which individuals produce info-bits themselves in an unbiased way. In the two remaining modes *select close info-bits* and *select distant info-bits,* a new random info-bit is created and presented to each individual analogously to the individual mode until the total number of info-bits is equal to the number of individuals. This can take some ticks, because info-bits are not always integrated by individuals. If the number of info-bits is equal to the number of individuals, each individual is presented a random existing info-bit which is inside (in the mode *select close info-bits*) or outside (in the mode *select distant info-bits*) a radius of size *D* around the individual's attitude position. This is done through the use of a recommendation algorithm that aims to present info-bits which the receiver will integrate with a probability higher than 0.5 (*select close info-bits*) or, respectively, an info-bit which confronts the individual with very different (but perhaps interesting) information. Thus, the number of potential new info-links per individual per tick is one in all of the modes, while the total number of new info-bits per tick varies.

2. **Individuals post info-bits to their friends.** If `posting` is activated, all individuals, one after the other in a random order, select a random info-bit from their memory and post it to all friends in their social network. All of their friends try to integrate the new info-bit. This represents the individual propagation of information through individuals in social media. When `posting` is switched on, an individual receives on average `numfriends` additional info-bits per tick. Note that these can be from friends with very different attitudes, because at this point there is no initial correlation between social network links and attitudinal proximity.

3. **Turn-over, refriending, and info-bit clean-up.** These settings serve as robustness tests and are thus not switched on in the baseline setting. Each individual dies with probability `birth-death-probability` (baseline 0) and is replaced by a new individual in a random position in the attitude space. This creates a certain fluctuation of individuals. Friend-links are created for the new individual such that the characteristics of the social network are preserved. New individuals start with no info-links.

Afterwards, each friend-link is subject to die with the probability `refriend-probability`. If a friendship is selected for potential death it stays alive for one more probabilistic event analogous to the integration of info-bits. Thus, friendships are more likely to vanish when the attitudinal distance of the friends is great, whereas friends with similar average attitudes are more likely to remain friends. When a friend-link dies, one randomly selected end of this link forms a new friend-link to a randomly selected friend of a friend with whom she is not friends yet. This refriend mechanism preserves the number of friendships.

Finally, each info-bit which is not held in any individual's memory is removed.

These three steps are repeated in the same order in every time step (i.e., tick).

Operationalization of the three levels of filters in the model

Different *individual* filtering mechanisms are represented by the parsimonious mechanism created to integrate new information based on the individual *latitude of acceptance*. The way it is modelled, this mechanism leads to information's being more likely retained within an individual's memory if it fits pre-existing attitudes. Thus, this individual filter represents cognitive biases, such as confirmation bias. This mechanism is active in all scenarios of the model, since such biases cannot be easily switched off. This is also true in reality. The mechanism's strength can be varied by changing the parameters of the latitude of acceptance.

Different *social* filtering mechanisms are represented by possibilities for individuals to post information to their network of friends, or to change their network of friends. The *info-posting mechanism* represents the function of real social media, where individuals post information to their friends (scenarios 3, 4, 7, 8). The *refriend probability* makes changes in individuals' social networks, in a way that friendships vanish when attitudinal distance is too great, and friends of friends may become new friends (Scenario 9).

Technological filtering (e.g., through recommender systems) is modelled by presenting individuals with info-bits that are either close to (in Scenarios 5 and 7) or distant (in Scenarios 6 and 8) from their average attitudes.

Additionally, to represent one-to-many communication vs. many-to-many communication, different modes of information-spreading (central propagation vs. social posting) are also part of the model.

Emergent outcome variables

We focus on the following macroscopic variables to measure filter bubbles, echo chambers, as well as fragmentation and clustering of individuals in attitude space.

- *Mean distance info-bits.* This is the average distance to the info-bits in each individual's memory. The average of these is the measure for the size of individual info filter bubbles.
- *Mean distance info-sharers.* Each individual has an average distance to all the other individuals with whom at least one info-bit is shared. The average of these differences

is the measure for the size of the filter bubbles of others with whom information is shared. Note, there need not be a friendship between information sharers.

- *Mean distance friends.* This is the average distance to all of an individual's friends. The average of these is a measure for the size of the individual filter bubbles of others to whom information is posted or who post information to the individual. Note, that friends need not be information sharers, because when their distance in attitude is large, they do not integrate information from the other.

While the mean info-bit distance is mainly determined by the latitude of acceptance *D*, the mean info-sharer and mean friend distance are more systemic properties which depend on emergent effects on the macro level.

Results

Initial settings of the model (baseline case)

We implemented the model in *NetLogo* (Wilensky, 1999) and ran simulation experiments, starting with our baseline configuration of 500 individuals from 4 communities, with on average 20 friends of which on average a fraction of 0.8 were from the same community. The birth-death and the refriend probabilities were set to zero in the baseline case (the latter is set to 1 in Scenario 9). The baseline parameters of the integration probability function were a latitude of acceptance of D=5 and a sharpness parameter of δ=20. That means that integration was very likely when the info-bit was at a distance smaller than five in attitude space and very unlikely otherwise.

When we initialised a simulation, individuals were all created at random positions in the attitude space. We ran simulations until the number of info-bits and info-links were saturated; this meant a stochastically stable situation had been reached, where all individuals were saturated with information (i.e., their memories were full) and there was little movement of individuals in attitude space. However, such a situation is often only a meta-stable one in the sense that structural changes are still possible in the extremely long run but with very low probability.

Overview

In the following section we will present the results of nine different scenarios of information propagation: In Scenario 1 we tested the effects of completely independent individuals, without any social mechanisms such as posting, who individually and randomly searched for and took in new information. In Scenario 2 individuals received central new info-bits, such as would have been the case with traditional mass media before any individualisation and without any social posting. In effect most people in society are influenced by the same central information in this way (*mainstreaming*; Griffin, 2012). In Scenario 3 (like Scenario 1, but with posting), individual searches for new information are combined with social posting. This set-up represents a society without any central information channels that would reach almost everyone. Instead, it is a society of individuals who randomly search for new information themselves and also post some of it to their network of friends. In Scenario 4, we tested effects of central news propagation which would reach almost everyone in combination with individualised social posting into one’s networks.

Scenarios 5 to 8 follow the same logic as 1 to 4, but with the addition of mechanisms for technological filtering through recommender systems. In Scenario 9 we reused Scenario 3 and added a refriend mechanism to test its impact.

Results for Scenarios 1 to 4

Figure 2 shows the positions of individuals in the attitude space and their info-links after stabilization. The info-bits are not shown in the figure because they would often cover the individuals. Their location can be assessed by the other empty end of the info-links. Friend-links and info-sharer-links are not shown for similar reasons. Nevertheless, they are part of the simulation. The quantitative characteristics are summarised in Table 1.

< Figure 2 >

< Table 1 >

In Scenarios 3 and 4 with social posting, the bubble of people with whom information was shared (as indicated by mean distance info-sharer) was smaller than the information bubble itself (as indicated by mean distance info-bits). This means that these individuals might have perceived strong attitude homogeneity with the people they shared information with and at the same time had the perception that they held diverse info-bits. This was different without social posting in Scenarios 1 and 2. Under the condition of pure individual info-bits (as in Scenario 1), there was no clustering and no info-sharing. Under the condition of central information propagation (Scenarios 2 and 4), there was some clustering, but the info-bridges between different clusters remained (individuals also shared information from other clusters). In our baseline case, the availability of social posting enforced strong clustering into typically four bubbles of info-sharers, who operated in slightly wider information bubbles with no informational contact to the other communities. These clusters evolved even though there was continual inflow and exposure to info-bits from the whole attitude space, because through social posting each individual became exposed to on average twenty info-bits from the social network, in addition to the one new info-bit from a random position. Info-bits from the social-networks were not randomly distributed, but were based upon the current distribution of info-bits. This implies that over time there would be a slightly higher concentration of people in a certain attitude region at the expense of other attitude regions which would lose agents. This concentration occurred, because the info-bits of agents in high concentration regions were propagated more than info-bits in regions with a low concentration of agents. The square geometry of the attitude space and the level of the latitude of acceptance then determined how many such concentrated regions ultimately emerge and remain.

A second result was that the friend network had no effect on attitude clustering. The bubble of friends maintained its large attitude radius, and clusters in no way self-sorted with respect to friendship communities. Figure 2 shows the communities of individuals by their colour. It is clearly visible that the info-sharer bubbles were composed of members of all four friendship communities. Actually, the formation of info-sharer bubbles would evolve as it would do with only one friendship community. Thus, attitude clustering was possible even though all individuals continuously received strongly differing info-bits from many of their friends, who held different average attitudes. They just did not integrate this information.

< Figure 3 >

Figure 3 shows exactly the same results as Figure 2, only the colours of individuals were set according to the results of a community detection algorithm (Louvain method; Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) that was applied to the network of info-sharer links. This shows that the eyeballed clusters were also detected by standard algorithms.

The impact of technological filters: Scenarios 5 to 8

We repeated the design above and analysed the new-info-modes *select close info-bits* and *select distant info-bits* without (Scenarios 5 and 6) and with social posting (Scenarios 7 and 8). This represents technological recommender systems which either provided individuals with information that was close to their average pre-existing attitudes (*close* info-bits in Scenarios 5 and 7) or that was distant from those attitudes (*distant* info-bits in Scenarios 6 and 8). Results are depicted in Figure 4 and their quantitative characteristics summarised in Table 1.

< Figure 4 >

The technological filter which selected close info-bits had an effect similar to social posting, even when posting was disabled. Individuals formed tight info-sharer bubbles with almost no shared info-bits between the bubbles. A filter which selected distant info-bits was able to sustain a fully connected info-sharer network. The simulation for this setting was much longer than for the other simulations, to test for very slow convergence into tight info-bubbles, which did not happen. Nevertheless, it might be possible that structural changes might possibly have happened after even longer time intervals. When social posting was switched on, the final outcome was very similar to the former scenarios with social posting: Four tight info-sharer bubbles evolved. The only difference is that it took much longer to reach meta-stability when only distant info-bits were selected by the technological filter.

The impact of refriending: Scenario 9

People not only map real-life friendship and family networks onto online social networks; they also defriend and refriend online contacts. The refriend mechanism in our model makes defriending more likely when the attitudinal distance is greater, while new friends are random friends of friends. In previous simulations, we had found that existing community structures in the friendship networks of information sharing with social posting led to info-sharer bubbles with people from all friendship communities (Scenarios 3 and 4). Scenario 9 in Figure 5 shows the situation after stabilization with the individual new-info-mode, and with social posting enabled as in Scenario 3, but additionally with a refriend-probability of one. This means that in each tick every friendship link was subject to removal when the attitudinal distance was too great.

< Figure 5 >

This result implies that echo chambers evolved when refriending happened in addition to social posting, but refriending was not the driving force for the formation of disconnected clusters of individuals with similar attitudes.

Larger latitudes of acceptance

We also tested the impact of larger latitudes of acceptance. In particular, we studied Scenario 2 (central new-info-mode without social posting) and Scenario 3 (individual new-info-mode with social posting) with a latitude of acceptance of 8 instead of 5. Figure 6 shows the results, which can probably also stand for all the other scenarios in which social posting was made possible.

< Figure 6 >

Interestingly, the larger latitude of acceptance led to a large consensual cluster with social posting, while much more diversity including some clustering remained with central information without social posting. This suggests that social media could also have the potential to bring about a societal consensus, which would not happen without but only with mass media with the same levels of acceptance. On the other hand, as we saw before, social media could also cause strong cohesive clusters maintained without any shared information with other parts of society through wider social networks.

An increase of the latitude of acceptance usually also implies that fluctuation of the attitudes of individuals increases. In the scenario with individual information and social posting individuals moved on average 0.04 per tick when the latitude of acceptance was at our baseline level of 5. For a latitude of acceptance of 8, the average individual move increased to 0.19. Thus, the societally beneficial consensual info-sharer bubble for the larger latitude of acceptance comes with the cost of larger individual uncertainty with lower attitude stability.

Discussion

Main findings

In Scenario 1 (without central information propagation, social posting, or recommender systems) no echo chambers emerged. Individuals spread out evenly in the attitude space (except for the extreme fringes). In Scenario 2 (with central information propagation, without social posting or recommender systems) distinct echo chambers emerged, but individuals still share some information with people outside their respective echo chambers. In Scenario 3 (without central info, with social posting and without recommender systems) distinct echo chambers without links between them emerged. This indicates strong attitude group polarisation. In Scenario 4 (with central info and social posting, without recommender systems) distinct echo chambers emerged as well.

Taken together, this shows that in our model filter bubbles and echo chambers evolved already from individual cognitive processes (modelled in all scenarios) in combination with central news sources that reach almost everyone, even without any social (posting or refriending) or technological (recommender systems) processes involved (Scenario 2). This implies that ‘traditional’ one-to-many communication would encourage the emergence of echo chambers. If, however, additional social posting processes occurred (simulating many-to-many communication; Scenarios 3 and 4), these echo chambers became more distinct and less interconnected, which would lead to even more fragmentation and polarisation of society.

In Scenarios 5 to 8 recommender systems were used to present new information to individuals. We found that recommendation of close info-bits had the same effect as social posting even without social posting, while recommendation of distant info-bits could maintain a connected info-sharer network. Social posting had the same effects as before.

The triple filter bubble framework

The triple filter bubble framework takes into account information filtering processes on the individual, the social, and the technology levels of analysis. It also indicates what effects these processes have on individuals as well as on society as a whole. While the filtering mechanisms on the different levels have been identified and described in previous research, their complex combination in a joint framework is novel. Results of our simulations show that the different filters interact and have effects on individual and social conceptual phenomena in ways that are not at all trivial. The complexity of the framework might seem like a disadvantage, since it makes it difficult to test the framework empirically with human subjects. However, our agent-based modelling approach allows for simulating all postulated processes (and others) step by step and simultaneously, and for gauging their joint effects.

Directions for future research

The results presented here are not a complete analysis of the behaviour of the model. Next steps of interest departing from our baseline case would be, for example, to study the effect of memory size or the degree of acceptance. Potential sensitivity analyses should concern themselves with the questions of whether the effects are similar in attitude spaces with one or multiple dimensions. We expect that the number of dimensions would have a strongly increasing effect on the number of evolving info-sharer bubbles. Furthermore, other distributions for the random appearance of info-bits should be tested. Additionally, the birth-death mechanism implies the possibility of adding additional dynamics to the network by having individuals leave it or join it and to explore how robust our findings would be with such a turnover (Kurahashi, Mäs, Lorenz, 2017).

Strengths and limits of agent-based modelling for social research

One strength of such simulations is that they facilitate empirical research in complex realms of reality that cannot be studied easily using experiments with real subjects. Additionally, phenomena can be studied that would be unethical to study with humans or that take a very long time to play out. This allows for new research questions and theoretical approaches to be empirically tested.

One possible critique concerns the old rule of thumb for simulations, *garbage in => garbage out*. The concern is that researchers could relatively easily create any desired model outcomes by trying out different rules and settings until their model fits their theory. To counter this critique, all model details should be made transparent, and the models used should be provided to other researchers to check, refute, or validate the initial findings. We hope to have shown that agent-based modelling represents a useful empirical approach to studying complex social phenomena and testing related theories or frameworks. Colleagues are invited to use our model [a preliminary version for peer review purposes can be downloaded at https://www.dropbox.com/s/m670o7rlq6g7ip9/triplefilterbubble_2.nlogo?dl=0], to (in-)

validate our findings, and to develop it further as means of answering additional research questions.

Conclusion

Modern technology cannot be stopped; people like to share their experiences digitally, and tech giants will further professionalise recommender systems to maximise the time users spend on their sites. On an individual level, these processes may lead to reassurance and enhancement of individually existing attitudes, behaviours, and identities. They increase individual attitudinal stability, and thus, individual certainty and security. On a societal level, however, these processes are prone to increase attitudinal differences between opinion groups and individuals and to cut communication ties between them, leading to attitude clusters, societal fragmentation, and polarisation. This poses a problem for modern democracies that rely on conflict resolution and reaching consensus through processes of democratic discourse. For the democratic process, it is necessary to be able to hear people express different opinions, to be willing to listen to them, and to engage in mutual discussion. So the digital world presents a genuine dilemma, where *positive* individual effects go along with *negative* societal effects. What can be done?

Generally, there are individual, social, technological, and societal solutions to these issues: On an individual level, knowledge about the processes leading to filter bubbles, or more generally, *media competence*, might mitigate these effects. On a social level, alternative mechanisms for debate, discussion, and creation of consensus are needed (possibly using social media). It is necessary to engage and discuss issues with people holding different opinions. On the technological level, means of increasing the serendipity of recommender systems are currently being discussed (Zhang, Séaghdha, Quercia, & Jambor, 2012) and will hopefully be implemented in the future.

On a societal level the deletion of *fake news* or unwanted content from the internet, i.e. censorship (as recently enforced in Germany for private companies like Facebook and the other tech giants), or the institutionalisation of the latter are proposed as solutions. However, since these measures limit the human right of free speech and damage free discourse, they may finally turn out to be more harmful than useful to a democratic society.

References

Abelson, R. P. (1964). Mathematical models of the distribution of attitudes under controversy. *Contributions to mathematical psychology*. New York: Holt, Rinehart & Winston, 142-160.

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review, 78*, 171-206. doi:10.1037/h0030834

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*, *70*(9), 1-70. doi:0.1037/h0093718

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130-1132. doi:10.1126/science.aaa1160

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012, April). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web* (pp. 519-528). ACM. doi:10.1145/2187836.2187907

Bazzan, A. L., & Klügl, F. (2014). A review on agent-based technology for traffic and transportation. *The Knowledge Engineering Review*, *29*(3), 375-403. doi:10.1017/S0269888913000118

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008. doi:10.1088/1742-5468/2008/10/P10008

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, *12*(4), 331-370. doi:10.1023/A:1021240730564

Castells, M. (2010). *End of millennium: The information age: economy, society, and culture.* Oxford: Wiley-Blackwell.

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems, 3*, 87-98. doi:10.1142/S0219525900000078

Fishbein, M. & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. *Reading, MA: Addison-Wesley*.

Fisher, S. & Lubin, A. (1958). Distance as a determinant of influence in a two-person serial interaction situation. *The Journal of Abnormal and Social Psychology, American Psychological Association*, *56*, 230-238. doi:10.1037/h0044609

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of Social Influence: Towards the Next Frontiers. *Journal of Artificial Societies & Social Simulation*, *20*(4), 1-31. doi: 10.18564/jasss.3521

Garimella, V., & Weber, I. (forthcoming). A Long-Term Analysis of Polarization on Twitter. https://arxiv.org/pdf/1703.02769.pdf

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, *14*(2), 265-285. doi:10.1111/j.1083-6101.2009.01440

Geschke, D. (2017). The RE-CO-KIT: A REality COnstruction KIT – technological, group dynamic, cognitive and motivational aspects. Poster presented at the *18th General Meeting of the European Association of Social Psychology*, Granada, Spain, July 5-8th 2017.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology, 78*(6), 1360-1380.

Griffin, E. (2012). *Communication Communication Communication*. New York: McGraw-Hill, 366–377.

Happer, C., & Philo, G. (2013). The role of the media in the construction of public belief and social change. *Journal of Social and Political Psychology, 1*, 321-336. doi:10.5964/jspp.v1i1.96

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5-53. doi:10.1145/963770.963772

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism Studies*, *13,* 815-824. doi:10.1080/1461670X.2012.664430

Hogg, M. A., & Turner, J. C. (1987). Social identity and conformity: A theory of referent informational influence. *Current issues in European social psychology*, *2*, 139-182. doi:10.1080/1461670X.2012.664430

Hosanagar, K., Fleder, D., Lee, D., & Buja, A. (2013). Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science, 60*(4), 805-823. doi:10.1287/mnsc.2013.1808

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design, 41*(4), 661-689. doi:10.1068/b120043p

Hunter, J. E., Danes, J. E. & Cohen, S. H. (1984). Mathematical Models of Attitude Change. *Academic Press*.

Krause, U. 2000. A discrete nonlinear and non-autonomous model of consensus formation. In: S.Elaydi, G.Ladas, J.Popenda, J.Rakowski (eds.), *Communications in Difference Equations*. Amsterdam: Gordon and Breach, 227-236.

Livingstone, S. (1996). On the continuing problems of media effects research. In J. Curran and M. Gurevitch (eds.), *Mass media and society, 2nd Edition*. London, UK: Edward Arnold, 305-324.

Loewenstein, George (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin, 116*(1), 75-98. doi:10.1037/0033-2909.116.1.75

Mason, W., Conrey, F. & Smith, E. (2007). Situating Social Influence Processes: Dynamic, Multidirectional Flows of Influence Within Social Networks. *Personality and Social Psychology Review, 11*, 279-300. doi:10.1177/1088868307301032

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review, SIAM, 45*, 167-256. doi:10.1137/S003614450342480

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2(2)*, 175-220. doi:10.1037/1089-2680.2.2.175

Nguyen, T. T., Hui, P. M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. ACM, 677-686. doi:10.1145/2566486.2568012

Noor, P. (2017). ‘The fact that we have access to so many different opinions is driving us to believe that we’re in information bubbles’: Poppy Noor meets Michal Kosinski, psychologist, data scientist and Professor at Stanford University. *The Psychologist, 30*, 44-47.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Penguin UK.

Resnick, P., Garrett, R. K., Kriplean, T., Munson, S. A., & Stroud, N. J. (2013). Bursting your (filter) bubble: strategies for promoting diverse exposure. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, 95-100. doi:10.1145/2441955.2441981

Sherif, M. & Hovland, C. (1961). *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change.* New Haven, CT: Yale University Press.

Smith, E. R. & Conrey, F. R. (2007). Agent-Based Modeling: A New Approach for Theory Building in Social Psychology. *Personality and Social Psychology Review, 11*, 87-104. doi:10.1177/1088868306294789

Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond.* Princeton, NJ: Princeton University Press.

Sunstein, C. R. (2009). *Republic. com 2.0*. Princeton, NJ: Princeton University Press.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of twitter. *Social Media+ Society*, *2*(3), 2056305116664221. doi:10.1177/2056305116664221

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences, 113*(3), 554-559. doi:10.1073/pnas.1517441113

Vinokur, A., & Burnstein, E. (1978). Novel argumentation and attitude change: The case of polarization following group discussion. *European Journal of Social Psychology, 8*(3), 335-348.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012, February). Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 13-22. doi:10.1145/2124295.2124300

Tables

Table 1: Summary of the quantitative characteristics of Scenarios 1 to 9

scenario #	mean distance info-bits	mean distance info-sharers	mean distance friends	No. of ticks until meta-stabilization[1]
1	3	NA[2]	15	700
2	3	4.1	14.6	600
3	2.6	0.8	11.2	570
4	2.8	0.9	12.7	570
5	2.8	1.4	11.3	1300
6	4.6	5.8	12.9	7500
7	2.6	0.9	10.6	570
8	2.5	0.7	12.2	2200
9	2.7	0.7	0.6	550

Notes: [1] These numbers mark the time step when the other values could be expected to stabilize at that level. It is not a precise measure or analysis of convergence time. [2] NA, because in this scenario shared info-bits could not exist.

Figures

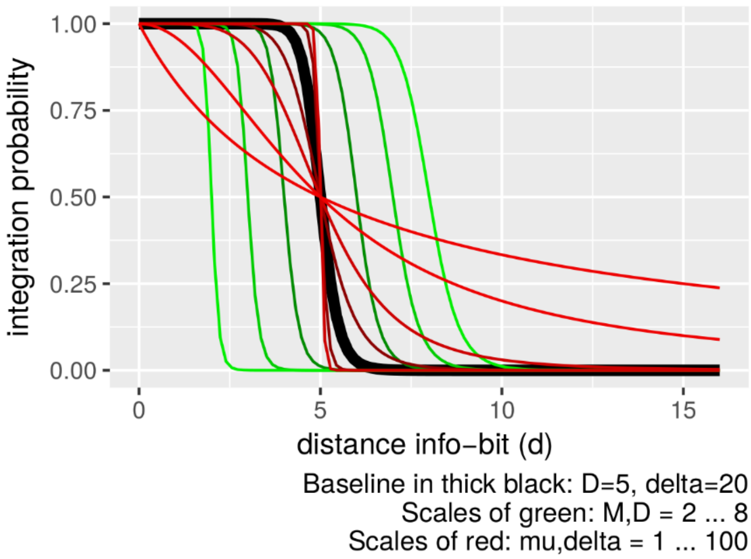

Figure 1: The functional form of $D^{\square}/(d^{\square} + D^{\square})$ from Equation (1). The fat black line marks the baseline case used for the simulation results in the following figures.

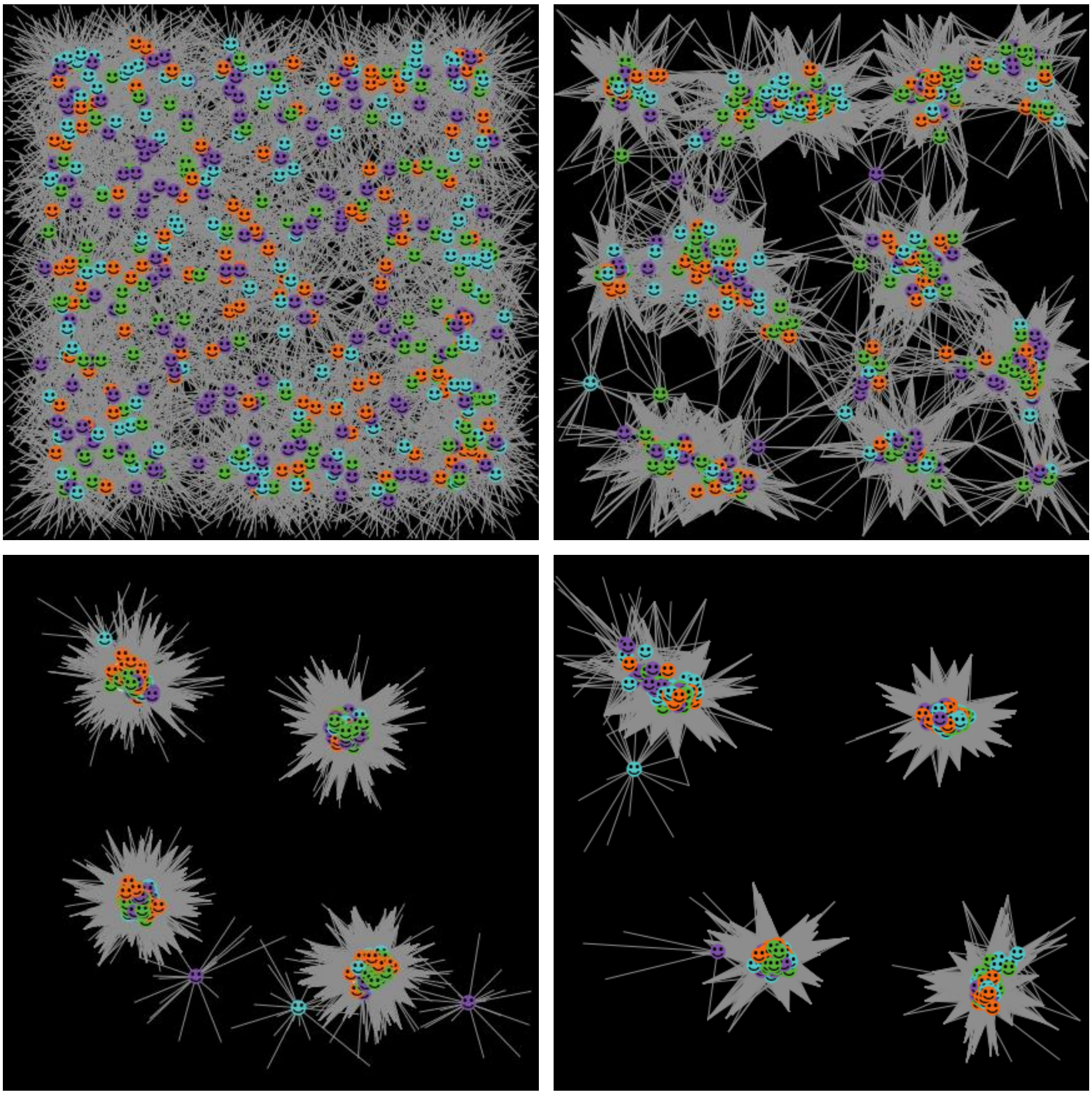

Figure 2: Individuals and their info-links after stabilisation for Scenarios 1, 2, 3, and 4 (from top left to bottom right row-wise). The colour of individuals determines their friendship community. On average, 80% of an individual's friends were of the same colour.

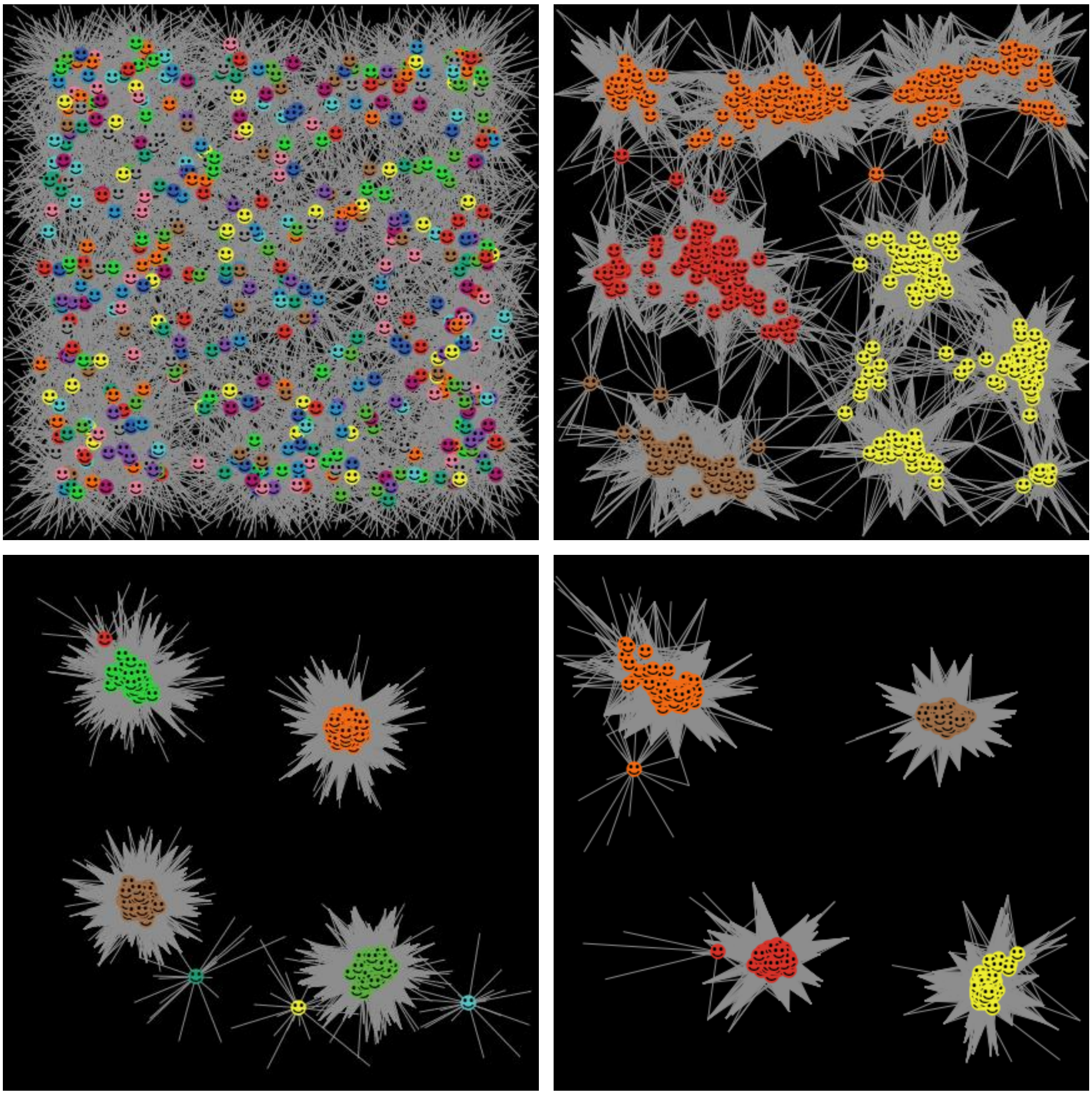

Figure 3: Same as Figure 2 with colours representing info-sharer communities computed by the Louvain method (Blondel, Guillaume., Lambiotte, & Lefebvre, 2008) in NetLogo's network extension.

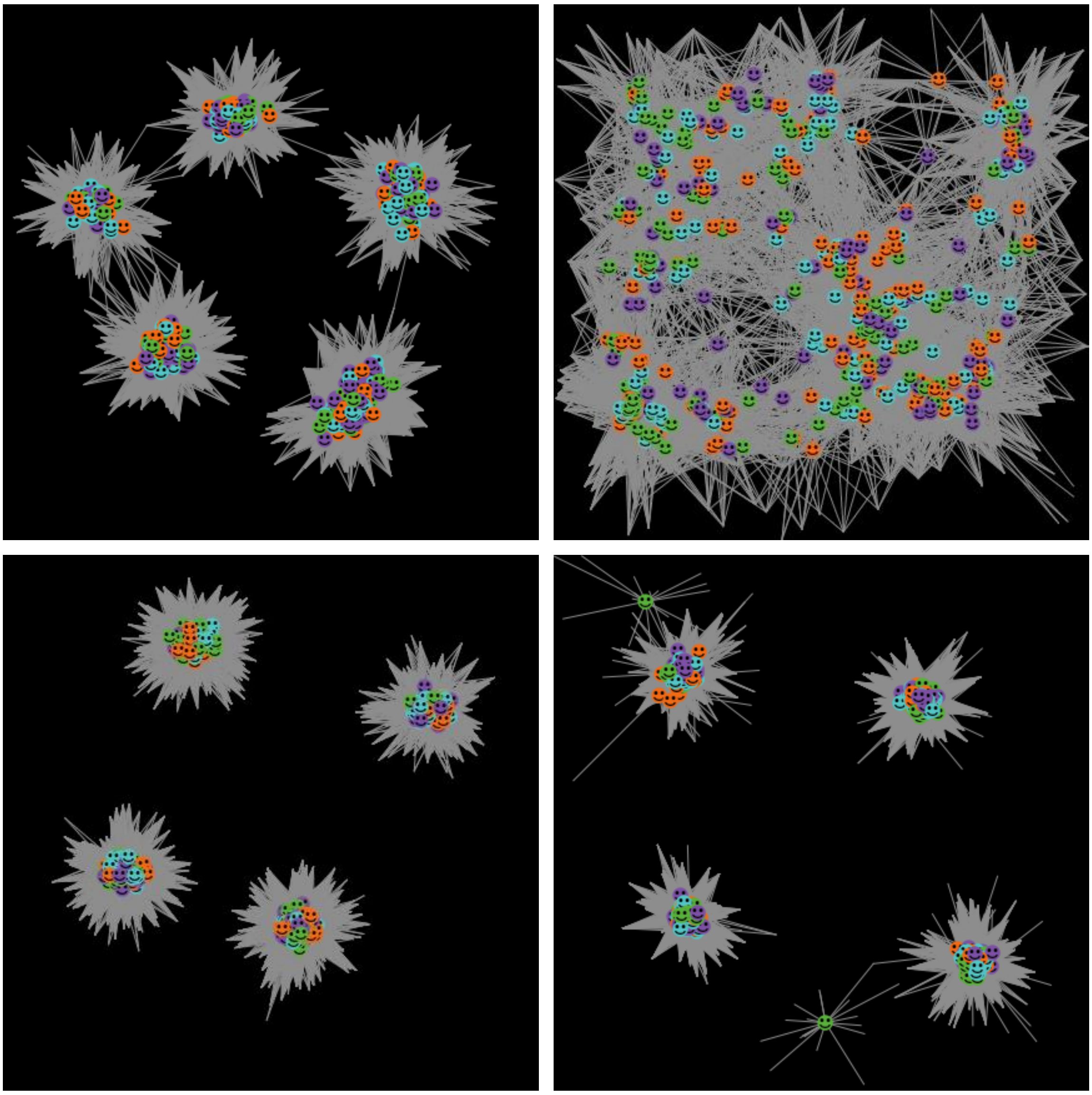

Figure 4: Individuals and their info-links after stabilisation for Scenarios 5, 6, 7, and 8 (from top left to bottom right row-wise).

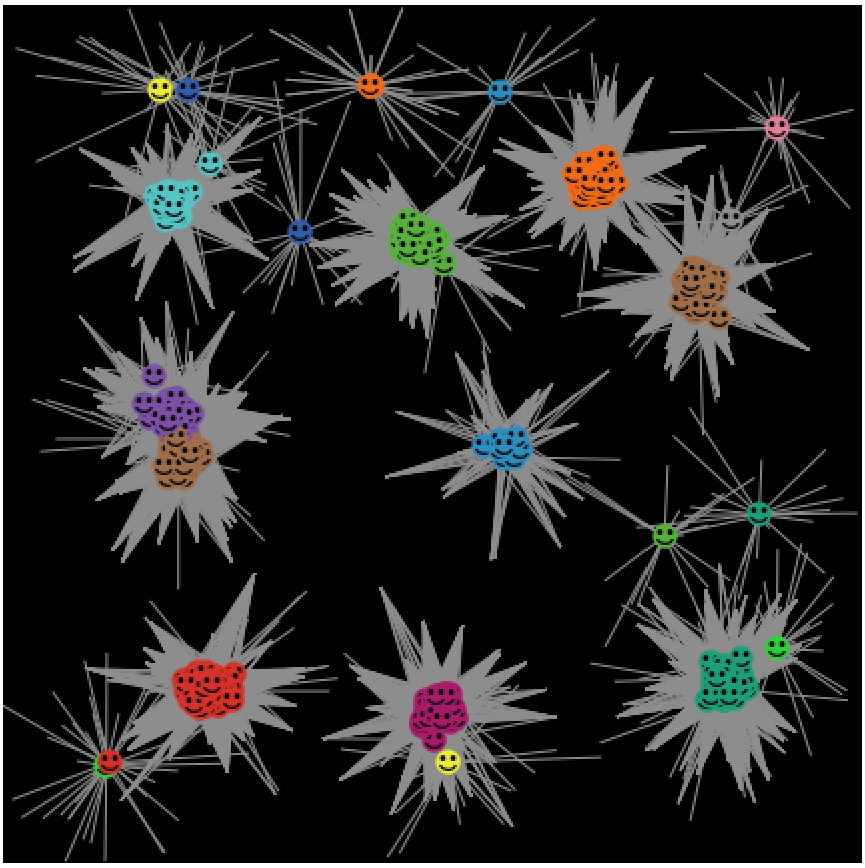

Figure 5: Scenario 9, individual new-info-mode with social posting and a refriend probability of one. Colours represent friendship communities computed by the Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) in NetLogo's network extension

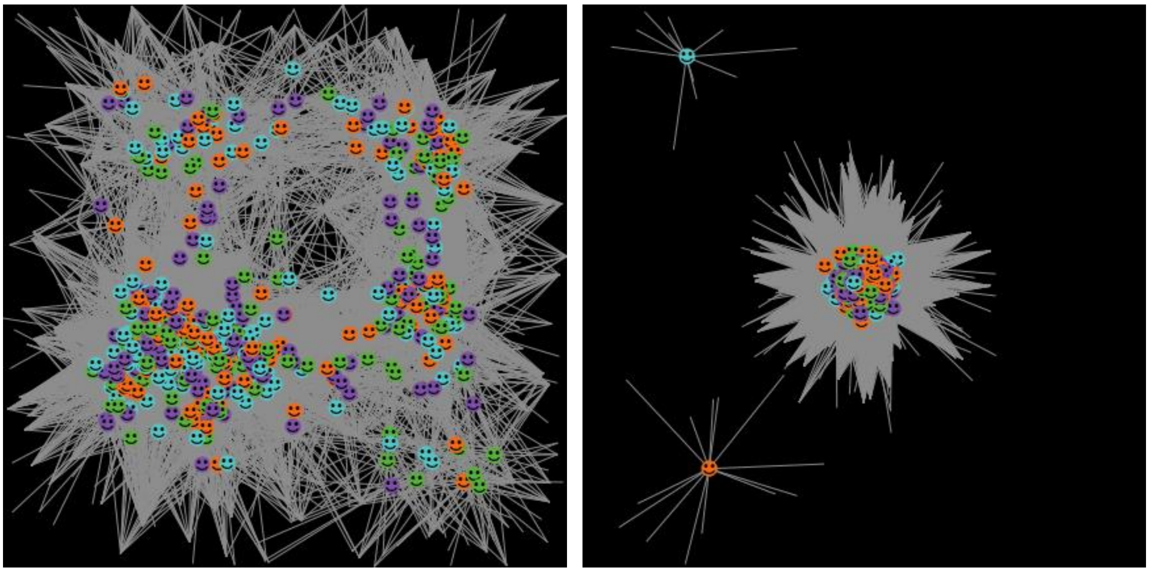

Figure 6: Scenarios with latitude of acceptance *D*=8 which is higher than the baseline case of *D*=5 used in all other scenarios. Left: Central new-info-mode without social posting (analogue to Scenario 2); right: individual new-info-mode with social posting (analogue to Scenario 3).