scientific reports

Check for updates

OPEN

The desire to avoid cognitive dissonance drives community formation in a social network model

Zoltán Kovács[1,2], Gergely Palla[1,3,4]✉ & Anna Zafeiris[1,2,4]

As the consequences of opinion polarization effect our everyday life in more and more aspect, the understanding of its origins and driving forces becomes increasingly important. Here we develop an agent-based network model with realistic human traits: individuals in our simulations are endowed with an internal belief system which they attempt to keep as coherent as possible. This desire – to reassure existing attitudes while avoiding cognitive dissonance – is one of the most influential and widely accepted theories in social psychology by now. Our model shows that even in networks that start out completely uniform (from a society of clones), this effort leads to fragmentation and polarization, reflected both by the individual beliefs (attitudes) and the emerging communities in the social network. By fine-tuning two parameters: (i) "dissonance penalty", measuring the strength with which agents attempt to avoid cognitive dissonance, and (ii) "triadic closure affinity", the parameter reflecting agents' likelihood to connect with friends of friends, a wide range of possible community structures are observed.

Keywords Social networks, Opinion dynamics, Polarization, Fragmentation, Cognitive dissonance, Belief systems

Network science provides a general approach for the study of complex systems on an extreme wide range, from chemical reactions within living cells through genetic regulation or neural interactions up to the scale of the Internet, the world trade or the global pandemic spreading of diseases[1–5]. One of the fields within network science of high interdisciplinary interest is the study and the modelling of the structure and dynamics of social networks, where nodes represent individuals and links indicate acquaintances or friendship relations. In general, the formation of social ties in such systems can be affected by many different factors such as the position and movement of people in the physical world (e.g., moving to a new town likely comes with the establishment of new connections) or the movement of people between institutions, firms, etc. (e.g., entering the university also helps finding new friends). However, a further key factor driving the formation of new ties and the deletion of existing links is the opinion or the beliefs of the individuals, with people usually preferring to be connected with others having similar beliefs[6,7].

By the mid-20th century, a new scientific discipline emerged focusing on exploring and simulating the process through which beliefs and opinions spread within human societies[8–11]. This area of study, rooted in statistical physics, mathematics and computer science[12–17], became known as *opinion dynamics*, and by now integrates aspects of various fields, such as psychology[18–22], biology[23,24] or political sciences[25,26].

Since the birth of this discipline in the mid-20th century, the suggested models become more and more refined and realistic regarding both the representation of agents and the social network by which they interact[12,27]. The first models highlighted some similarities between the Ising model and the way people align their opinions[28] (agents' opinions being the spins which are pushed to be aligned with their neighbours'). Some other early models suggested similar framework, but with continuous opinions (taking values from a certain, pre-defined interval, usually between 0 and 1 or -1 and 1) instead of binary ($+1/-1$, yes/no, etc) states[16,29,30]. In general,

[1]Department of Biological Physics, Eötvös University, Pázmány P. stny 1/A, Budapest 1117, Hungary. [2]MTA-ELTE Lendület "Momentum" Innovation Research Group, Eötvös University, Institute of Archaeological Sciences, Budapest 1088, Hungary. [3]Health Services Management Training Centre, Semmelweis University, Kútvölgyi út 1., Budapest 1125, Hungary. [4]Gergely Palla and Anna Zafeiris contributed equally to this work. ✉email: gergely.palla@emk.semmelweis.hu

these models predict global consensus[31] and polarization appears only due to some extra stipulations, such as threshold of communication[32] or distancing[33]. Later, in a seminal paper, direct psychological properties have been also suggested to be incorporated into the models[34], making a big step towards more realistic models. Other approaches proposed to extend the one-scalar representation of agents into multi-dimensional vectors[35] which has motivated new directions, among others, studies related to the emerging and co-evolving social networks[36]. Even more recently, models have been proposed in which the elements of the belief systems are interrelated[37–41] while being embedded into a social environment as well [42,43].

In the present work we follow the common approach of combining networks with agent-based simulations: We use two types of networks. A social network, where the nodes represent agents and the links reflect the social ties between them, while another network represents the inner state of their belief system. In this second aspect we depart from other models, by introducing a biologically and psychologically more realistic representation for the belief system of the agent, which is equivalent to a (relatively small) weighted network between nodes corresponding to concepts, as illustrated in Fig. 1a. The interactions in the outer social network with other agents can modify the connection weights in this inner representation of the belief system of the agent. This is inspired by a process called "associative learning", referring to the observation that the fundamental learning process in humans – and in animals as well – is based on creating new associations between already existing concepts (which process, of course, can be supplemented by incorporating new concepts as well)[44–46]. A further important driving force in our model is *cognitive dissonance avoidance*, a widely accepted theory in cognitive psychology[47,48], by which agents modify their attitudes towards various concepts and beliefs – represented by signed node-weights in the inner state representation – in a way that their beliefs and attitudes remain as coherent as possible (contradiction-free). Lastly, similarly to the majority of evolving social network models, the probability for the formation of new links in the (outer) social network between the agents as well as the probability for dispatching an existing link are affected by the similarity between the attitudes of the agents[6,7].

The novelty of our approach lies not in discussing the above-mentioned human characteristics, rather in demonstrating their fundamental effect both on our attitudes and on the structure of social networks, and, importantly, the way they can be integrated into agent-based models. We are able to show that an initially random social network between "clones" (agents with completely identical internal belief system) – depending on the actual set of parameters – may facilitate the emergence of a rich variety of human group structures, including consensus, fragmentation and polarization. In addition, we also find that the "repulsive interaction" effect (a.k.a "distancing", when agents' opinion further depart in case of a certain amount of disagreement[49]) is a natural sequel of the effort of keeping the coherence level as high as possible, and as such, there is no need to introduce this rule "by hand"[50].

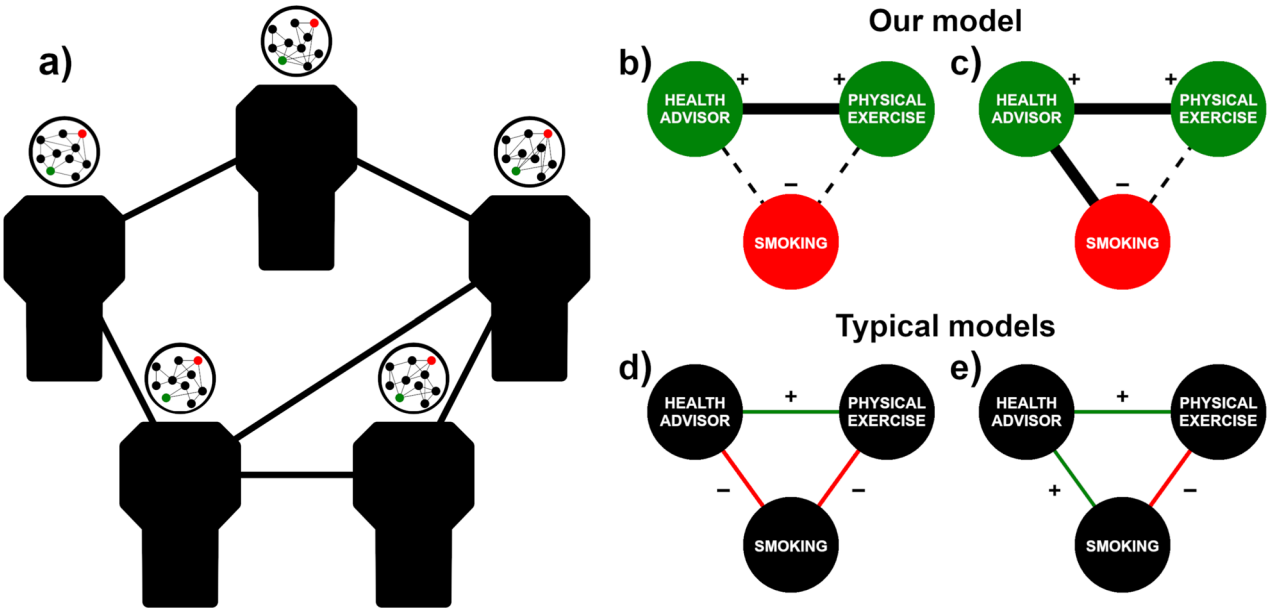


Fig. 1. Social network of agents with internal belief systems. **a**) Illustration of the social network of the agents where the belief systems of the agents are represented by additional internal networks. The coloured nodes correspond to the concepts with fixed attitudes, one having a constant +1 value (green) and the other having a constant -1 value (red). **b**) A 'coherent' triangle of concepts in our model, where two concepts with the same signed attitude (positive in this case) are strongly associated while the association between concepts with opposite signed attitudes are weak. **c**) An 'incoherent' triangle in our model, where a strong association appeared between two concepts having opposite signed attitudes. **d**) A 'coherent' triangle of concepts in models which operate with signed connections between the concepts. **e**) An 'incoherent' triangle in models based on signed connections in the belief system.

For readers approaching the model from psychology or the social sciences, we note that our framework can be viewed as a minimal formalization of three widely studied cognitive mechanisms: (i) associative learning, where concepts that co-occur in thought or communication become more strongly linked; (ii) cognitive dissonance reduction, where people adjust their evaluations to maintain coherence among strongly associated concepts; and (iii) homophily-based tie formation, where individuals tend to maintain social connections with others whose expressed attitudes preserve internal coherence. The aim is not to replicate the full complexity of human cognition, but to distill these well-established principles into a tractable agent-based model that allows us to explore how their interaction shapes emergent social phenomena such as consensus, polarization, and fragmentation.

Model description

Our approach considers an evolving social network, where the nodes are agents. This network is directed, where the out-links of an agent indicate the communication probabilities with other agents. We assume that each agent has an internal state (what we can refer to as a belief system) that affects the communication with the out-neighbours and also the establishment of new connections. Furthermore, we also assume that the communication acts affect the internal state of the recipient agents, thus, in parallel with the social network structure, the internal belief systems of the agents also change over time.

This internal belief system is modelled by a network (graph) in which the nodes are *beliefs* or *concepts* (which we use as synonyms in the present paper), node *values* – a scalar value ranging from -1 to $+1$ – represent the agents' *attitudes* towards these concepts with negative values indicating negative sentiments and positive values showing support, and finally, the *links* of the network reflect the *strength of association* between the given nodes (concepts).

This representation – where attitudes are encoded as real numbers ranging from -1 to $+1$ – follows a widely used convention in classical opinion dynamics models, such as the Deffuant bounded confidence model [32], the Hegselmann-Krause model [51], and related frameworks reviewed in Ref. [12], in which numerical values denote the degree of support (positive) or rejection (negative) of a concept or opinion.

In our model, associations – which are the *links* in the internal belief systems – represent the *strength of cognitive co-activation* between concepts (nodes), manifesting the Classic Hebbian principles ("neurons that fire together wire together")[52] and other well-established distributional-learning models[53–55], all of which argue that the mere co-occurrence strengthens the internal link between the two concepts, regardless of the attitudes themselves (values of the nodes). (For a more detailed examination of association networks, an interactive tool presenting measured data in various languages is provided in Ref. [56]).

Regarding the emergence of cognitive dissonance, consider for example a person who values a healthy lifestyle. In this belief system, "health advisor" is likely evaluated positively, while "smoking" is evaluated negatively. If the individual learns that the health advisor is a smoker, a new association is formed between a positively evaluated concept and a negatively evaluated one, leading to dissonance. In contrast, learning that the health advisor is a proponent of physical exercise links two positively evaluated nodes, resulting in reassurance – the inverse of dissonance – as captured by Eq. 1, detailing how coherence is calculated in our belief system.

Accordingly, there are two types of networks being studied in parallel in our simulations: (i) a social network, where nodes represent agents and links indicate the likelihood of communication, and (ii) belief networks, where nodes represent concepts or beliefs and links correspond to the association strength between these beliefs. Each agent, naturally, possesses a unique belief system.

Overview of the main concepts

The most important traits we incorporate into our model are the following:

1. *Internal state represented by a weighted network*. In humans, opinions and beliefs are never isolated; every idea or belief is interconnected with others. As a matter of fact, humans cannot even remember something unless it is associated with something meaningful[57]. In other words, concepts and beliefs are arranged into an organized framework, often referred to as a *belief system*. According to that, in our model the opinion or internal state of the agents is not a scalar or even a vector, instead it is represented by a small internal network between the beliefs or concepts, as illustrated in Fig. 1a. In this approach, the non-negative link weights (falling into the $[0, 1]$ interval) represent the strength of association between two concepts, whereas signed node values (falling into the $[-1, 1]$ interval) indicate the agents attitude towards the given concept or belief, with negative values indicating condemnation and positive values referring to support. A more detailed figure showing both the social network and the individual belief systems from a simulation can be found in the Supplementary Materials.
2. *Communication is enhancing the association between the topics.* New information often comes in the form of connecting originally unrelated concepts, and these new associations may immediately entail the re-evaluation of the pre-existing beliefs and attitudes. In accordance with that, in our model, information, coming through social interactions change the connection weights between beliefs in the internal belief system of the interacting agents. More concretely, during a communication act, the initiator agent is choosing a pair of concepts (serving as the "topic" of the conversation) according to a probability proportional to their connecting weight in its belief system, and as a result of the communication act, the association strength (link-weight) of the same link is enhanced also in the belief system of the recipient agent. However, in order to prevent all association strengths in the belief system approaching 1, after we modify them they are re-normalised so that the sum of the total association strengths in the belief system remains constant. From a psychological perspective, the normalisation step reflects the well-documented limits of human associative capacity. People do form new associations when concepts co-occur, yet strengthening one association necessarily reduces

the relative salience and accessibility of others, because attention, working memory, and cognitive resources are finite. In other words, individuals cannot indefinitely increase the strength of all associations at once; increased focus on one conceptual link naturally diminishes the prominence of others. The constrained-sum normalisation therefore serves as a simple formal representation of these capacity limits, ensuring that the internal belief network remains psychologically realistic by preventing indiscriminate saturation of all associations.

3. *Cognitive dissonance avoidance.* A key characteristic of human belief systems is our strong desire to maintain coherence between our beliefs, while striving to steer clear of the uncomfortable sensation known as cognitive dissonance, which is basically the feeling of *in*coherence. Based on that, one of the main driving forces of our model is that after communication acts, the recipient agents try to minimise the cognitive dissonance of their internal belief system by modifying their attitudes towards the concepts, that is, the node values in the internal state representation. To illustrate how cognitive dissonance may arise in our approach, let us imagine someone having a *health advisor* recommending *physical exercise*, which are connected to each other by a strong link, since they are related (associated) to each other, as shown in Fig. 1b. In contrast, *smoking* is unrelated to physical exercise (since they are activities that people typically do not pursue in parallel), so the link weight between them is zero or close to zero. Similarly, the link weight between the *health advisor* and smoking is also zero, because they are not associated either. (Fig. 1b). In contrast, if our agent gains knowledge that the health advisor is a smoker, the link between them becomes strong, as shown in Fig. 1c. Let us also assume that the agent has a positive attitude towards physical exercise and the health advisor (green colour) and negative attitude towards smoking (red colour) in these examples. In the case of Fig. 1b, concepts with the same sign are connected by a strong connection, whereas concepts with opposing sings are not connected (or are only very weakly connected), thus, we have coherence. In contrast, in the case of Fig. 1c, the strong connection between smoking and the health advisor, having opposite signed node values, leads to in-coherence, inducing cognitive dissonance. According to the above, the *coherence* of a single pair of concepts in the belief system can be defined simply as the product of the (signed) attitudes and the strength of the association between them, whereas the coherence level of the entire belief system can be given as the sum over the pairs of concepts. For comparison, in Fig. 1d-e we show the traditional approach (which is motivated by the Balance Theory[58,59]) for representing coherence and in-coherence among concepts via signed connection weights. Here, by multiplying the signs of the three edges appearing in a triad[41,60,61] we may end up with a positive or a negative result, former indicating a stable, coherent triad and the latter corresponding to an unstable, in-coherent one.
4. *Universally positive and negative concepts.* Beside cognitive dissonance avoidance, another fundamental feature of the human mind is that we share some universally positive concepts which remain positive independently of our education or socialization. For example, due to its deep biological and psychological roots, the concept of *mother* is typically associated with unconditional love, care and protection, and as such, forms an unconditional, invariably positive concept[62–64]. In contrast, for example *fear* or *death* are uniformly negative concepts[65,66]. Furthermore, other concepts might also tend to be "constant" positive or negative, depending on the culture the agent belongs to[67]. That is, some other constant negative or positive attitudes might differ from culture to culture. In our model, such concepts can be incorporated by nodes in the belief system that have an unalterable weight of either $+1$ or -1. For sake of simplicity, in the present work we have incorporated a single concept with constant positive weight of 1 and another single concept with constant negative weight of -1. These two concepts are the same for all the agents, depicted by the green and the red colour in Fig. 1a. The rationale for incorporating fixed positive and negative concepts stems from psychological and evolutionary theory. For example, in Ref. [68], it is argued that humans are uniquely aware of their mortality, which creates persistent existential anxiety. Cultures – through religion, ritual, and belief systems – evolve in part to manage this fear. This perspective – ever since widely demonstrated empirically as well [69] – shows that the awareness of death has widespread influence on moral behavior, creativity, and social identity. Consequently, *death* is a prime candidate for a universally negative, fixed node in the belief network. Conversely, several studies identify universal positive concepts. For example, in Ref. [70] core moral foundations like Care/Harm, Loyalty/Betrayal, and Sanctity/Degradation are associated with deeply positive values across cultures. For instance, concepts such as "mother," "child," "family," and "life" typically carry strong, unchanging positive valence. These insights support our decision to fix a small number of positive and negative concepts in all agents' belief networks.
5. *Expanding the social network with the friends of friends (TCA function).* In real-life social systems, individuals continuously reshape their social networks: connections fade or strengthen, and new ties are formed as people acquire new interests, change environments, or meet others via shared acquaintances. One of the most well-established empirical patterns in this process is the tendency for "friends of friends" to become friends themselves–a phenomenon known as *triadic closure*[71–73]. This pattern can be understood as a consequence of two reinforcing mechanisms. First, according to the principle of homophily–well-documented in social network research–people are more likely to form social ties with others who are similar to themselves in terms of attitudes, beliefs, or background characteristics[7]. If person A is similar to person B, and B is similar to person C, then A and C are also likely to share similarities, making them predisposed to interaction. Second, the structural opportunity for contact increases when two individuals share a common friend: they are more likely to encounter each other at the same social events or be introduced through mutual contacts, increasing the probability of interaction[73]. Our model captures this realistic feature through the triadic closure affinity (TCA) mechanism, which modulates the likelihood of agents interacting with their friends' friends. We justify this addition not only based on empirical observations but also due to its theoretical importance in network evolution models.

In our approach the time evolution of the social network is manifested in the change of the connection weights between the agents, where each social link can take a value in [0, 1]. After a communication act initiated by agent *i* with an other agent *j*, the weight of the out-link on *i* towards *j*, denoted by w_{ij} is increased or decreased based on the similarity between the attitudes of the two agents with regard to the topic of the communication (a pair of concepts). In parallel, the weight of the out-link on the recipient node *j* towards *i* is also updated, however here we use the after-stabilization attitude, with the updated attitude values in the agent's belief system.

Detailed model description

Our model features *N* agents, each with their own belief system containing *M* nodes (concepts). These concepts might range from political positions to football teams and public figures, among others. Each agent has an attitude towards these concepts, represented by a node value which is quantified on a scale from -1 to $+1$, indicating completely negative to completely positive views, respectively. The un-directed connections between concepts are also weighted with link weights ranging from 0 to 1 , which signifies the strength of the association between the concepts. For instance, if an agent learns that their favorite football team is implicated in a match-fixing scandal, the association between these two concepts would increase.

Based on their attitudes, two concepts can either reinforce each other (if their sign is the same), leading to *reassurance*, or conflict with one another (one being positive, the other negative), causing *cognitive dissonance*. For instance, if a politician whom a person favors supports a policy that the individual also endorses, linking the politician with the policy causes reassurance (it increases the level of coherence). Conversely, if that same politician is associated with a disfavoured political agenda, then this association causes cognitive dissonance. We assume that reducing cognitive dissonance is more crucial than merely achieving reassurance (tuned by the parameter $d > 1$)[74]. Accordingly, for a pair of concepts α and β with respective node values (attitudes) a_α and a_β, the impact on the coherence level of the agent is defined as

$$C_{\alpha\beta} = \begin{cases} d \cdot a_\alpha \cdot a_\beta \cdot B_{\alpha\beta}, & \text{if } a_\alpha \cdot a_\beta < 0 \text{ (dissonant attitudes)} \\ a_\alpha \cdot a_\beta \cdot B_{\alpha\beta}, & \text{otherwise (aligned attitudes)}, \end{cases} \tag{1}$$

where *d* is the dissonance penalty parameter and $B_{\alpha\beta}$ indicates the association strength between α and β in the belief system. Although in real-life societies *d* probably follows a distribution, in the present model – for sake of simplicity – we handle it as a constant parameter.

The total coherence level of the agent can be calculated by summing over all concept pairs as

$$C_{\text{tot}} = \frac{2}{M(M-1)} \sum_{\alpha=1}^{M} \sum_{\beta=\alpha+1}^{M} C_{\alpha\beta}, \tag{2}$$

where we have divided the sum by the total number of concept pairs in the belief system. In this way, a fully coherent belief system has a coherence of $C_{\text{tot}} = 1$. As we shall detail later, the agents try to maximize the coherence level of their belief system after each communication act in which they took part as a 'receiver'.

The *N* agents are all part of a directed and weighted social network, where link weights can range between 0 and 1. The outgoing links of the agents represent the willingness that they choose to communicate with the other agents. The social network evolves through this communication between individuals. (A flowchart of the social network simulation is provided in Fig. S1 in the Supplementary Information.)

At the beginning of each iteration we choose uniformly at random the agent who will be the initiator of the communication act. The probability that the chosen initiator agent *i* will communicate with an other agent *j* is depending mainly on the weight of the link pointing from *i* to *j*, denoted by w_{ij}. Furthermore, we also incorporate *triadic closure*[72,75] into the model, corresponding to a widely used concept in network science, referring to an increased likelihood for the formation of triangles, resulting from connecting to a friend of a friend. With formula, the probability for communicating with *j* is proportional to

$$\hat{p}_{ij} = w_{ij} + \frac{\sum_{q \neq i} w_{iq} w_{qj}}{TCA(\sum_q w_{iq})}, \tag{3}$$

where *TCA* is a (monotonously increasing) *triadic closure affinity* function depending on the out-strength (sum of the out-weights) of *i* (see Eqs. 8, 9 and 10). The role of *TCA* is to ensure that triadic closure has a stronger effect on the communication of agents with a low strength value and a weaker effect on 'hubs' that have a high number of strong out-links. The actual probability for choosing *j* as the recipient in the communication act is simply given by normalising the $\hat{p}_{ij}$ values as

$$p_{ij} = \frac{\hat{p}_{ij}}{\sum_j \hat{p}_{ij}}. \tag{4}$$

In the case where an agent is completely isolated ($\sum_j w_{ij} = 0$), the communication partner will be chosen randomly from all other agents with uniform probability.

Once the recipient agent is fixed, a pair of concepts α and β are chosen to serve as the 'topics' of the communication. The probability for choosing (α, β) is proportional to the weight $B_{\alpha\beta}$ between α and β in the belief system of the initiator agent *i*. That is, the topic choice depends on the association-strength, not attitude value. As the result of the communication, the weight of the connection between the same two concepts is

increased by a random number between 0 and 1 in the belief system of the recipient agent *j*. (Naturally the new $B_{\alpha\beta}$ in agent *j*'s belief system is capped at 1 if an overshoot should occur). After that, all weights in the belief system of *j* are normalised to keep the total sum of association strengths constant through the process. (A flowchart of the communication process is provided in Fig. S2 in the Supplementary Information).

To account for the desire to maintain a coherent belief system, the recipient agent *j* is also permitted to adjust the node values within its belief system in response to the updated $B_{\alpha\beta}$ values. For instance, if agent *j* seeks to reduce cognitive dissonance between their favourite football team and the concept of match-fixing, they might choose to either diminish their positive attitude towards the team or enhance their acceptance of match-fixing. In our model, this is captured through a stabilisation process, where the agent may attempt to randomly alter its node values (attitudes) over *b* iterations. The details of a single iteration in the stabilisation process can be listed as follows:

1. Choose a uniform random number in the $[-1, 1]$ interval.
2. If this random number has the same sign as the attitude to be stabilised, scale the random number by the distance of the attitude from the corresponding extreme pole.
3. Add the random number to the attitude.
4. Recalculate coherence with the updated attitude value using equation (2)
5. Accept the new attitude value with the following probability:

$$p = \exp\left(\frac{C_{new} - C_{orig}}{T}\right). \tag{5}$$

Here C_{new} and C_{orig} are the new and original coherence values, and T is a temperature parameter.

The recipient agent performs the above stabilisation process, first on each of the two concepts discussed during the communication for $b - b$ iterations, then for another b iterations where a concept is chosen at random in each iteration. In total, the number of stabilisation process iterations performed is $3b$. A flowchart of the exact stabilisation process can be seen in Fig. S3 in the Supplementary Information.

We note that the above rules for the update of the association weights between the concepts and the attitudes at communication acts inherently modify the belief system even when a pair of identical agents in full agreement with each other interacts. For example, it is possible that the increase of the association weight between the 'topics' of the communication together with renormalisation of the aggregated association weights (slightly reducing all the other weights) may decrease the coherence for the recipient agent. As a result, this agent may change its attitudes towards some of the concepts, introducing disagreement between the initially identical agents.

Naturally, the communication also affects the links in the social network. For the initiator node *i*, this is based only on the attitudes of the beliefs discussed during the communication. Therefore, the update rule for the weight of the out-link from *i* to *j* is defined as

$$w_{ij}(t) = w_{ij}(t-1) + (a_\alpha^{(i)} \cdot a_\alpha^{(j)} + a_\beta^{(i)} \cdot a_\beta^{(j)}) \cdot r, \tag{6}$$

where *r* is a random number between 0 and 1, $a_\alpha^{(i)}$ denotes the attitude of the *i* regarding concept α, etc. In parallel, the recipient agent *j* will also change the link-weight of its out-link pointing towards the initiator agent *i*, but with the new attitudes after the stabilization process, given as

$$w_{ji}(t) = w_{ji}(t-1) + (a_\alpha^{(i)} \cdot a_\alpha^{(j)\text{new}} + a_\beta^{(i)} \cdot a_\beta^{(j)\text{new}}) \cdot r. \tag{7}$$

In the case of Eqs.(6-7), if the new weights would fall outside of the $[0, 1]$ interval, they are set to 0 or 1, respectively.

The two most important parameters in our simulations were:

(1) The dissonance penalty, *d*, is the parameter appearing in (1) that controls the strength with which an agent can or cannot tolerate cognitive dissonance, with higher values referring to smaller tolerance. In other words, it reflects the strength by which an agent is "bothered" by experiencing cognitive dissonance.
(2) The triadic closure affinity, *TCA*, is a function that appears in (3), controlling the likelihood with which agents choose to communicate with a 'friend of a friend' in the social network, instead of their own neighbours.

We study 3 cases:

$$TCA(x) = x \tag{8}$$

- This is the least strict case, that is, the most open for communication with friends of frien ds.

$$TCA(x) = x^2 \tag{9}$$

- This is a stricter case, where the number of friends and the normalizing factors do not scale linearly. This means that agents communicate less through their friends' social networks if they have enough people in their own circles. For example, if an agent has only 2 friends, the relative probability of communicating with one of their friends is only 1/4 compared to the agent communicating through their own connections. If they have 10 friends, this probability further decreases to 1/256.

$$TCA(x) = x^4 \tag{10}$$

- This is the strictest case, where the number of friends reduces the willingness to communicate outside of direct friends even more. An expanded analysis of the triadic closure functions and an example social network with no triadic closure effect can be found in the Supplementary Material Section S5.

Another important assumption in our model is the existence of fixed-value concepts. To explore the role of these more rigorously, we also conducted simulations in which no beliefs had fixed positive or negative attitudes. In these simulations, the system consistently evolved toward a state where all attitudes reached an extremum – either all $+1$ or all -1. While this might initially appear to be a modeling artifact, we offer a more meaningful interpretation: in the absence of any universally "anchored" values – such as fear, pain, or threat on one end, or life and growth on the other – the internal drive for coherence naturally favors uniformity. This uniformity can emerge as either utopian optimism (all positive) or nihilistic pessimism (all negative), depending on initial fluctuations and social reinforcement. This dual attractor phenomenon aligns with existential and evolutionary psychological theories that view negative attitudes as adaptive responses to threat, and positive attitudes as signals of progress toward meaningful goals.

Simulation results

Our model consisted of 100 agents in a fully connected social network with all link-weights being equal to $w_{ij} = 1$ at the beginning of each trial. The agents were all clones (identical copies) of an initially generated agent, sharing the exact same belief system, both association strengths and attitudes. The belief system size was $M = 10$, having one concept with constant positive attitude $a_\alpha = 1$ and also another concept with constant negative attitude $a_\beta = -1$. The temperature parameter was set to $T = 0.01$.

For each trial, we set the parameters and ran the simulation until it converged and reached its final state, where the main structural properties of the social network were no longer changing. The precise criteria of convergence are described in the Methods. We note however that links are constantly formed and deleted in our model, thus, these end states were static only in the statistical sense. The results below refer to these steady, final states, and contain the results of 100 runs for each data point. To interpret our results, we used a wide variety of measures, of which a brief overview can be found in table 1.

Polarisation and fragmentation of the social network

Our general observation is that the defined social network model can reach a wide variety of stable end states, including consensus (where basically all agents have a more or less similar opinion and are willing to communicate with each other), polarisation (where two large groups are formed with a clear opinion difference) and fragmentation (where the network falls apart into small isolated components). We note that although the social network always starts from a fully connected network with unit link weights, during the time evolution of the system whenever a link weight w_{ij} becomes zero, that link can be considered to be removed from the network. As an illustration of the wide range of possible behaviours of our model, in Fig. 2 we show the layouts of three networks, corresponding to the stable end states of simulations at different parameter settings.

The final states are shaped by the dissonance penalty *d* and the *TCA* function. With a low dissonance penalty (referring to high tolerance towards cognitive dissonance), agents are less likely to alter their attitudes, resulting in no change within the social network since the agents are identical to each other (Fig. 2a). However, when the dissonance penalty is higher, agents may deviate from their initial set of attitudes.

Regarding the *TCA* parameter, which determines the probability of interacting with friends of friends according to Eqs.(8-10), we observed that when using a simple linear function $f(x) = x$, at higher values of the

Measure	Meaning
t-SNE and UMAP	Visual representation of the attitude space
Attitude homogeneity	Average normalized Manhattan distance of non-constant attitudes
Extremism	Average absolute attitude
Sum of link weights	Total existing link weights inside the social network
Number of communities	Number of communities detected by the Leiden algorithm
Largest component ratio	Fraction of nodes inside the largest connected component out of all nodes
Association strength distribution	Statistical distribution of all association strengths inside the agents' belief systems
Belief system homogeneity	Average normalized Manhattan distance of links in the belief system

Table 1. Measures and their meanings used in the results. We use a wide variety of measures capturing different aspects of our model. This table provides an overview for easier interpretation.

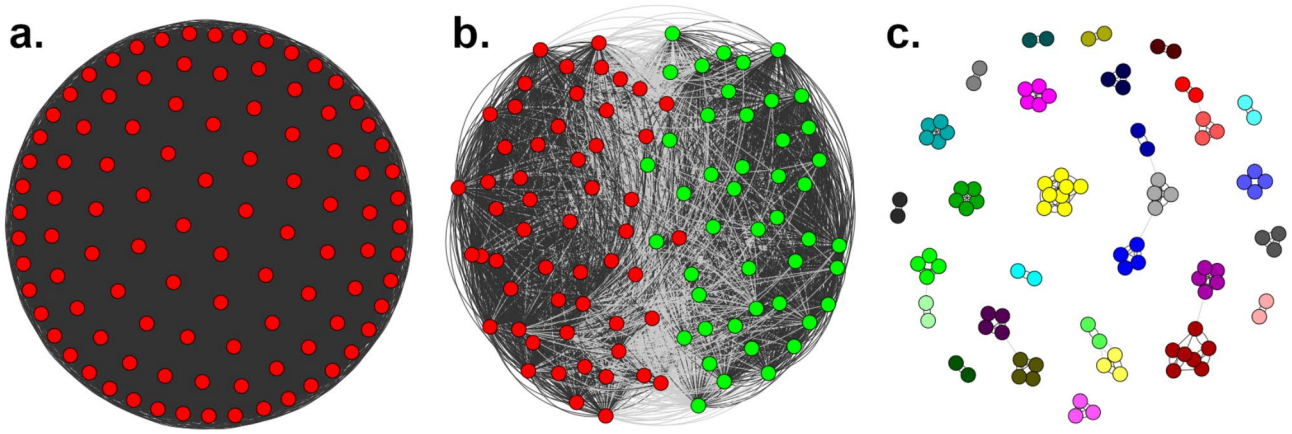


Fig. 2. Stable end-states of the social network. Three different states of the social network after convergence in three simulations with different parameter settings. **a)** With $d = 1$ and $TCA = x$ consensus emerges. **b)**$d = 3.5$ and $TCA = x$ results polarisation, where two communities form, indicated by colours (extracted using the Leiden algorithm [76]). **c)** When using $d = 3.5$ and $TCA = x^4$, we can observe fragmentation of the network into small isolated components.

dissonance parameter d the social network generally divides into two communities (Fig. 2b). These correspond to dense sub-networks that are more loosely connected to each other. Communities, also called as modules, groups or clusters are known to occur not only in social networks but also in various other types of networks, and community finding is in general an intensively researched topic in network science [77–79]. In the present work, we used the Leiden algorithm [76] to identify communities, a highly efficient method known for producing high-quality partitions of input networks based on modularity [80] (corresponding to the most widely used metric for evaluating community quality). Both the Leiden algorithm and the modularity are briefly described in Methods.

When nonlinear *TCA* functions such as $f(x) = x^2$ are applied, the average number of groups increases, typically ranging between 3 and 4. If the nonlinear effect is made even stronger with a function like $f(x) = x^4$, we can see the groups fragmenting further, as agents become increasingly unlikely to expand their social circle once they have established their own friends. In such cases, the social network in the end state is composed of numerous small isolated components (Fig. 2c).

Divergence of the attitudes

In all simulations, our agents were initially identical clones, sharing the same association strengths and attitude values. However, as the interaction progressed, their belief systems began to diverge from the initial state. Especially at larger d values (corresponding to lower tolerance levels towards cognitive dissonance), at some point during the simulation, the occurrence of a drastic change was typical during which the agents divided into two (or more) communities (groups, clusters, etc.) with respect to both their belief systems and the structure of the social network. This is illustrated in Fig. 3, comparing the initial state of the system with a later stage after a certain amount of communication.

To visualize the similarities or differences between the belief systems of the agents, we employed t-distributed Stochastic Neighbor Embedding[81] (TSNE), as shown in the left column, and Uniform Manifold Approximation and Projection[82] (UMAP), displayed in the middle column. Both methods use the attitudes (node values) of the agents as input vectors and provide a nonlinear projection of this data into two dimensions. (A brief description of these embedding methods is provided in the Methods section). As illustrated in the top row of Fig. 3, at the initial state where the agents' belief systems are identical, they form very tight clusters in the attitude embedding spaces, with minimal variance occurring solely due to the stochastic nature of the embedding algorithms.

In contrast, after certain amount of communication, the cloud of agents splits into two clusters according to both TSNE and UMAP, as indicated in the bottom row if Fig. 3. Simultaneously, the structure of the social network has also transformed (right column of Fig. 3), evolving from an initially fully connected network into a network comprising two dense communities. These communities (also called modules or groups) were located with the Leiden algorithm[76]. (A brief description of the Leiden approach is given in Methods). The colouring of the agents in the left and middle panels reflect their community membership in the social network. As it can be seen, there are agents who are on one side of the attitude space but still belong to the opposing social group. These are those agents who have already changed their attitudes but haven't communicated enough yet to leave their current community.

In order to quantify the similarities and differences between the attitudes of the agents at the group level, we define a quantity we call *attitude homogeneity*, $AH \in [0, 1]$, calculated from pairwise comparisons of agents' attitudes and relying on the absolute difference between the node values. (The precise definition of AH is given in Methods). This metric takes a value of 1 if, and only if all considered agents have exactly the same attitudes. In parallel, we also define *extremism*, E corresponding to the average of the absolute magnitude of the node values in the belief system. (A larger E values indicates that the given agent has more attitudes falling closer to the extreme $+1$ or -1 values).

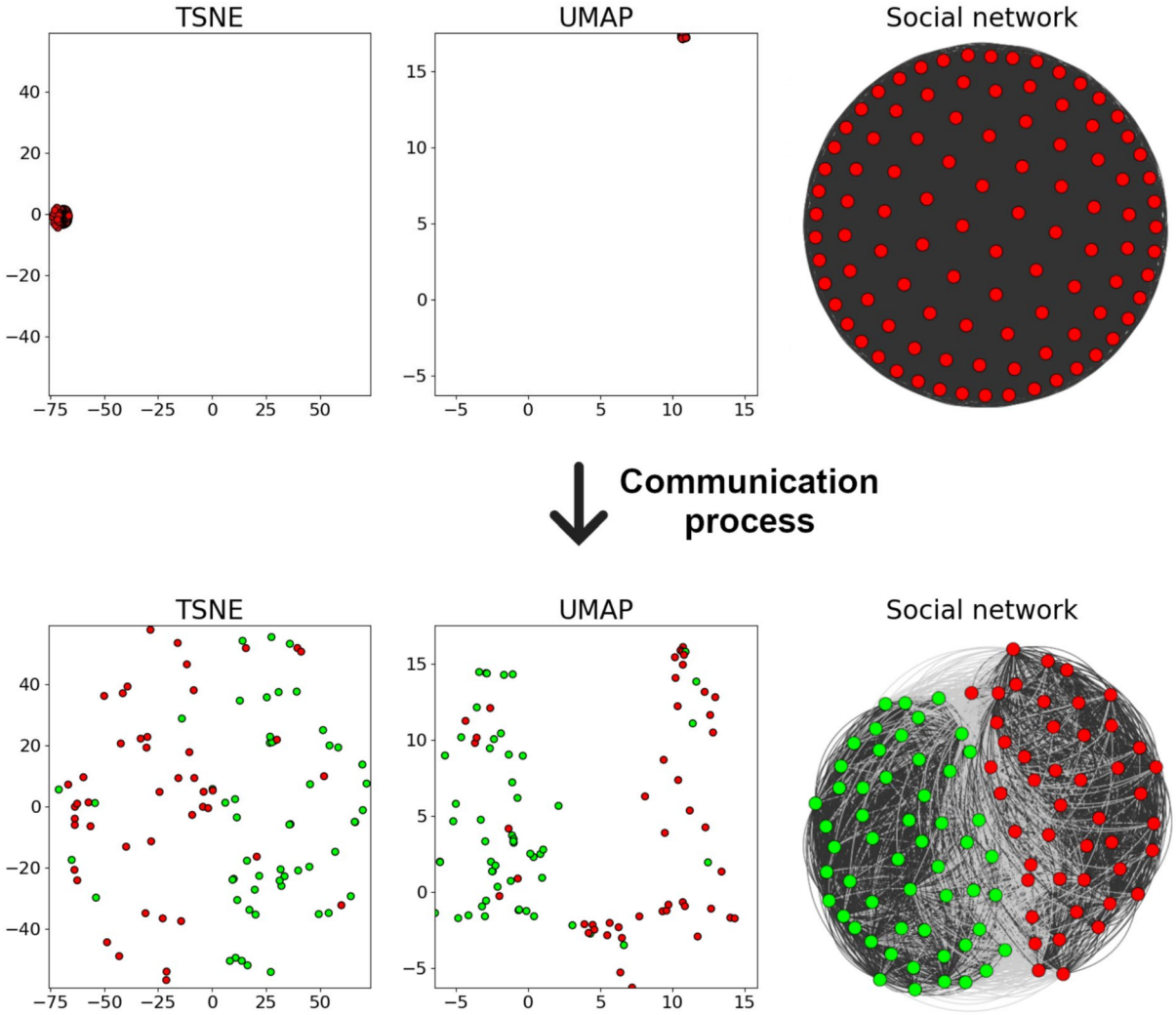


Fig. 3. Attitude embeddings and social network after communication process. The top row shows the embedding of the attitudes (node values in the belief systems) according to the TSNE (left) and UMAP methods with the social network displayed on the right for the initial state of the simulation. Here the social network is fully connected and is consisting of agents with identical belief systems. At this state, the clones form a dense cluster in the attitude space, with only minor differences due to the stochastic nature of the TSNE/UMAP embeddings. In the bottom row we show similar results after a certain amount of communication, where the social network forms two densely connected communities (marked by the different colours). These communities can also be seen in both embeddings, roughly corresponding to a left/right split.

In Fig. 4 we show the attitude homogeneity AH and extremism E at the end state of the simulations as functions of the dissonance penalty d for the three different TCA functions.

According to that, the attitude homogeneity of the entire network (shown in blue) drops sharply around a dissonance penalty value of $d = 2$. In other words, above a certain level of *in*tolerance towards cognitive dissonance, attitudes begin to rapidly diverge. As we shall show later, this is also the point where we can observe an important change in the network structure of the end states as well, where for low d values we usually observe one single community and for $d > 2$ two or more communities emerge. Therefore, when multiple communities form in the social network, members belonging to the same communities develop similar attitudes that are at the same time different from the attitudes of the agents in other communities. Accordingly, the homogeneity of the communities (shown in red) lacks this sharp drop and shows a slower decrease as a function of d.

The intuitive explanation for the above behavior derives from the avoidance of cognitive dissonance, which acts as a fundamental drive in our model. Typically, agents experience the most coherent state when their attitudes are close to one of the two extreme values, 0 or 1. Nonetheless, as the penalty for cognitive dissonance (d) increases, a new optimal point emerges around 0, marking an "indifferent" stance by which the agent avoids cognitive dissonance (as well as reassurance). This state is generally temporary, with agents often reverting to one of the extreme attitudes, possibly even switching to the opposite compared to their initial attitudes. Consequently, repeating the interaction process may lead the group to adopt a full spectrum of attitudes, with an equal distribution of agents at both extremes.

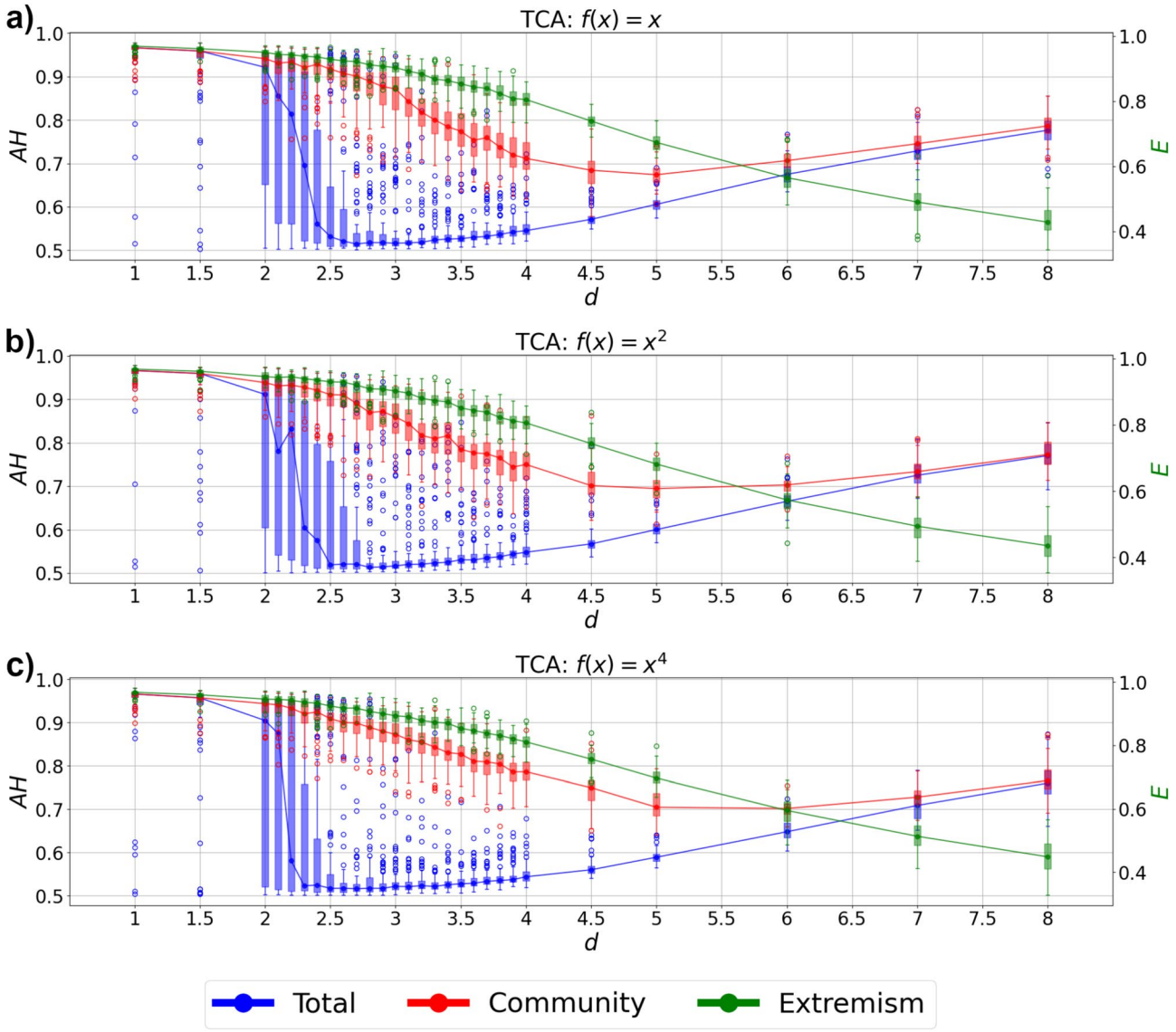


Fig. 4. Attitude homogeneity (AH) and extremism (E) values. Boxplot of the *AH* and *E* over 100 runs as a function of the dissonance penalty *d*. The *AH* calculated for the entire network is shown in blue, whereas the *AH* evaluated inside communities (found by the Leiden algorithm) is shown in red. In parallel, the extremism is also plotted in green using the right vertical scale. The median values are connected by continuous curves. The three panels correspond to different *TCA* functions as indicated by the panel titles. The total homogeneity values begin to diverge when approaching $d = 3$, while the community homogeneity remains high. As extremism starts to drop, around $d = 6$ the majority of agents form a moderate core, which decreases the differences between attitudes, and the total and community homogeneities become similar again.

However, if we increase *d* further, we can reach a point where this "neutral" stance becomes stable, in which case the attitudes of all agents will stay more similar overall. This represents the formation of a "moderate" majority, who are neutral on almost all issues. They are usually flanked by a smaller number of extremists on both sides.

The above explanation is also supported by the behaviour of the extremism values (shown in green), which are monotonically decreasing as a function of *d* in Fig. 4. This indicates that the increase of homogeneity in the large *d* regime is accompanied by more and more agents taking up "neutral" or less extreme attitudes.

We include a more thorough analysis of the individual agents behaviour at different dissonance penalties in the Supplementary Material of this article, in Section S3. We also found that the divergence of attitudes occurs in a fixed social network as well, for which a short example is included in Section S6 of the Supplementary Material.

Structural changes in the social network

In parallel with the changes in the belief systems of the agents, the social network can also undergo major structural reorganisation in our model. In Fig. 5 we show the total sum of the link-weights, $W = \sum_{ij} w_{ij}$ in

the end state of the simulations as a function of the dissonance penalty, d for different *TCA* functions. According to that, at low d values, roughly up to $d \simeq 1.5$, the social network seems to retain its fully connected nature, where the sum of the link weights remain close to the possible maximal value. However, around $d \simeq 2$, for all considered *TCA* functions a sudden drop occurs in the sum of the link weights, indicating a drastic change in the network structure. According to Fig. 5, for the $f(x) = x$ case around half of the possible links remain after this drop, but for the two stricter *TCA*-s, a vast majority of the links disappear, resulting in a sparse social network.

In the large d regime an increase in the aggregated link-weights can be observed for all studied *TCA* functions. However, for the most strict $f(x) = x^4$ case (Fig. 5 bottom panel) this increase is only very mild, whereas for the other two *TCA* functions (Fig. 5 top and middle panels) the sum of the aggregated link weights can grow back close to its initial value when d becomes very high. This is because at these d values, most agents have close to neutral attitudes. When two such agents communicate, as we can see from equations 6 and 7, the change in the weight of the social tie is minimal. Agents will still communicate through these links with a relatively high probability, and it would require multiple such communication attempts for the tie to completely disappear.

The drop in aggregated link weights is accompanied by the emergence of communities in the network structure. As mentioned earlier, in the present study we used the Leiden algorithm for community detection. In Fig. 6 we display the average number of communities found by the Leiden method in the end state of the simulations as a function of d (blue symbols and curves). For all studied *TCA* functions, at low d values (i.e.,

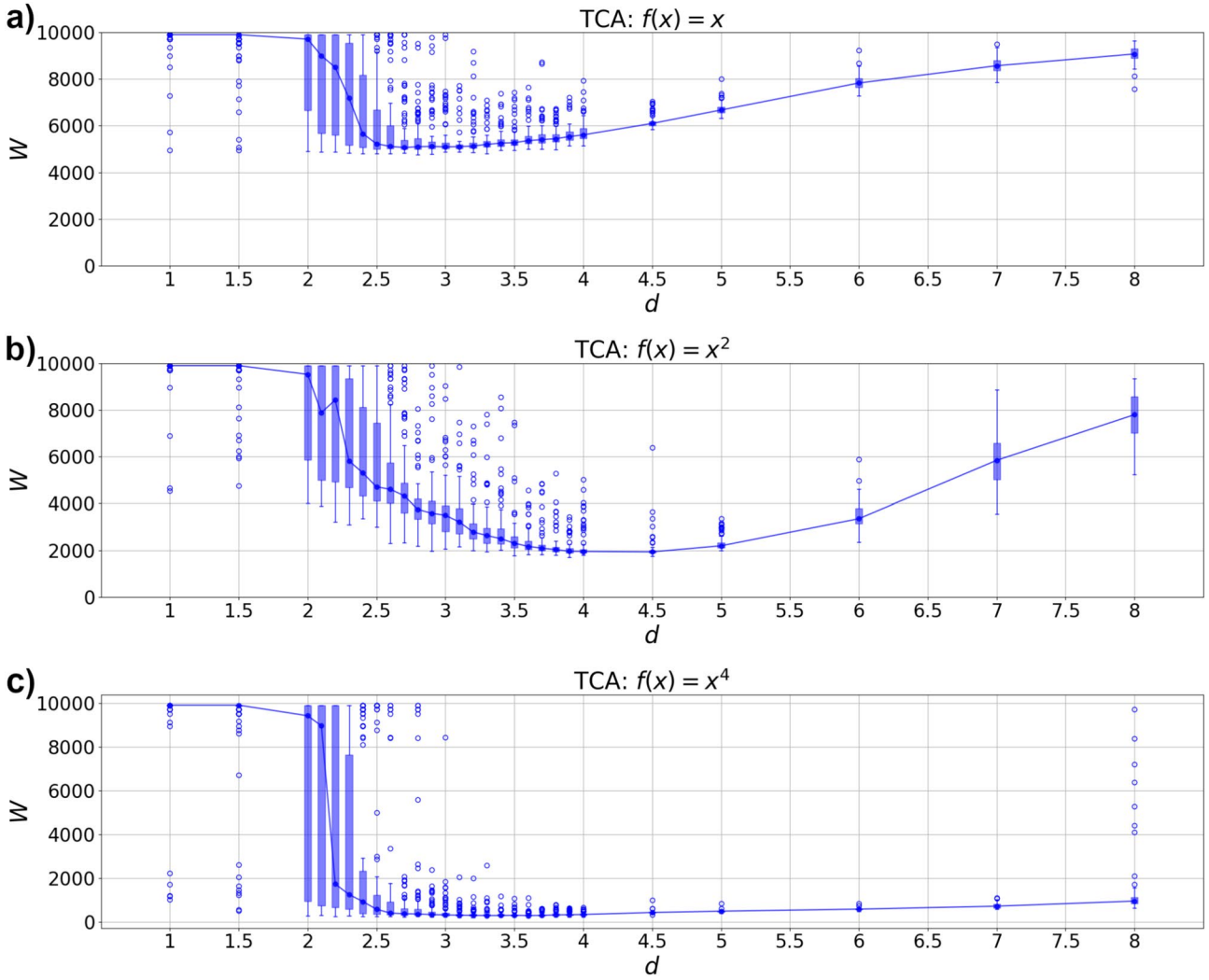


Fig. 5. Sum of the link weights in the social network. We show the aggregated link weights, given by $W = \sum_{ij} w_{ij}$ after convergence, as a function of the dissonance penalty d. The symbols represent a standard box plot over 100 runs for each parameter setting according to d, and the median values are connected by a continuous curve. The three panels correspond to different *TCA* functions as indicated in the panel titles. Close to $d = 2$, W starts to decrease rapidly in all cases, indicating the disappearance of social links in the network. However, depending on the *TCA* function, the remaining number of links can vary from around half of all links to a few hundred. For higher d values, this effect decreases, and the end states remain more similar to the original fully connected network.

below roughly $d \simeq 1.5$) the social network always consisted of a single community. This changes at the indicated transition point, above which for the $f(x) = x$ *TCA* function (Fig. 6 top panel), in most cases the network splits into two communities, marking polarization. For the more strict *TCA* function of $f(x) = x^2$ (Fig. 6 middle panel) the average number of communities is increasing up to roughly 4-5 communities in the larger d regime. In contrast, for the most strict *TCA* function of $f(x) = x^4$ (Fig. 6 bottom panel) the social network breaks into a larger number of small communities, consisting of only 2-6 agents each for moderate d values.

In Fig. 6 we also show the relative size of the largest connected component (red symbols and curve). Based on that, for the *TCA* functions of $f(x) = x$ and $f(x) = x^2$ (Fig. 6 top and middle panels) the social network always consists of a single component, meaning that there is still communication between different communities. In contrast, for the *TCA* function $f(x) = x^4$ (Fig. 6 bottom panel) the communities may also completely detach from the rest of the system, as indicated by the decrease of the relative size of the largest connected component roughly around $d = 3$. In other words, for this choice of the *TCA* function at moderate d parameter values we can observe the fragmentation of the network into small isolated components that do not communicate with each other. However, when we choose higher d values (that is, roughly above $d = 5$), the relative size of the largest

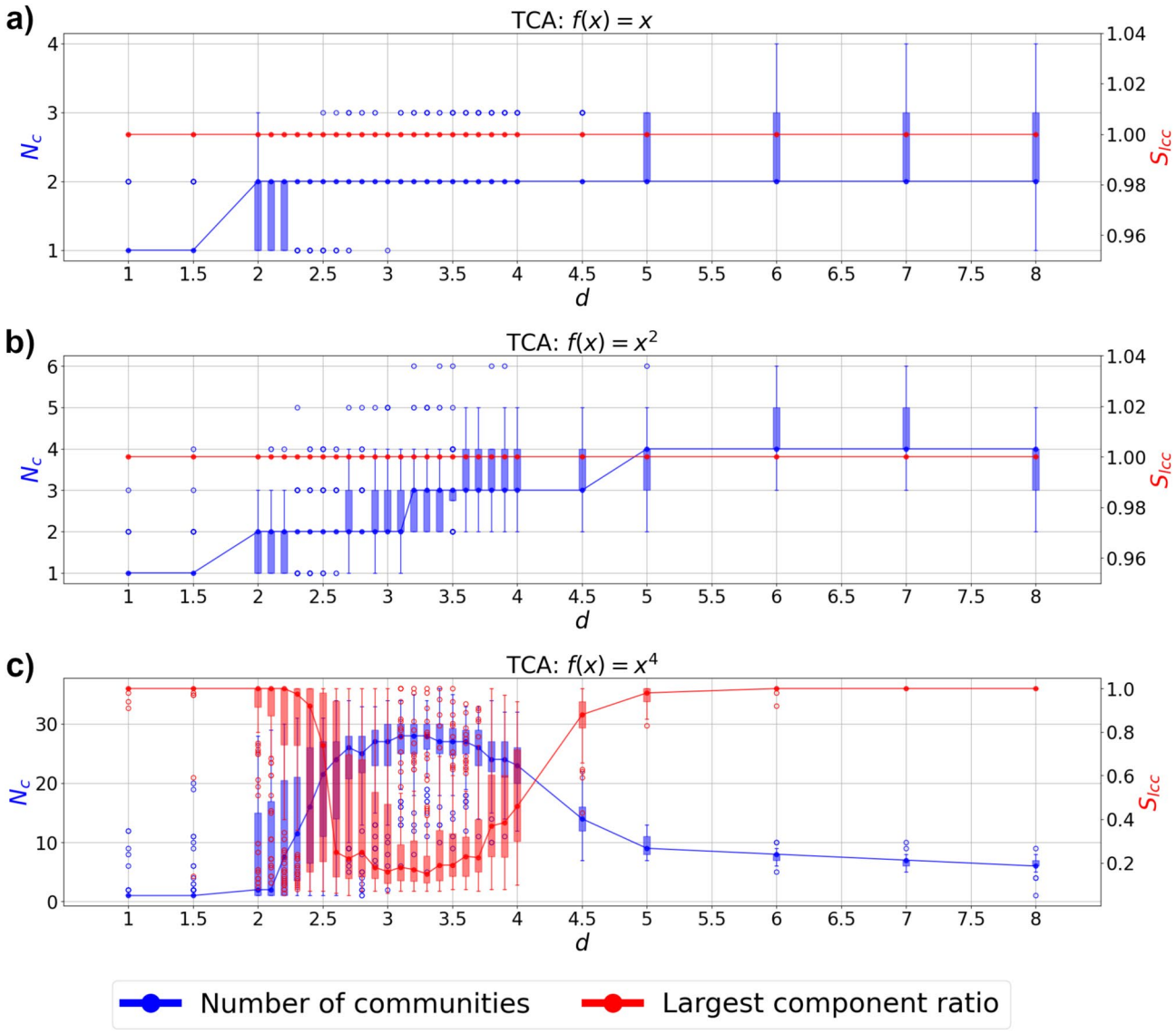


Fig. 6. The average number of communities N_c and the relative size of the largest connected component $S_{lcc} = N_{lcc}/N$. The number of communities detected by the Leiden algorithm is shown on the left vertical axis as a function of the dissonance parameter, d. The relative size of the largest connected component compared to the total number of agents is shown on the right vertical axis. The symbols represent a standard box plot over 100 runs for each parameter setting according to d, and the median values are connected by continuous curves. The applied *TCA* function is indicated in the panel titles. In all three cases, the number of communities starts to increase at $d = 2$, and reaches its maxima around $d = 3$. For TCA: $f(x) = x^4$ case, this also results in the breakup of the social network into multiple components. At $d = 5$, this effect starts to disappear.

connected component increases back to one, meaning that the network does not fall apart into disconnected components. This is accompanied by a decrease in the average number of communities, indicating that instead of the small isolated modules agents are instead organised into larger communities. This is because at higher dissonance penalties, the extremism of attitudes decreases as shown in Fig. 4. The outcome is an increase in agents with neutral attitudes, reducing the incentive to destroy social ties. Consequently, the social network remains more densely connected.

Changes in the structure of the belief systems

The formation of communities in the social network entangled with the development of clusters in the attitude embedding space raises the question whether we can observe related structural changes also in the belief systems of the agents. In order to investigate that, in Fig. 7 we show the distribution of the link weights in the belief systems (referring to the strength of associations between concepts and beliefs) averaged over all the agents in the end-state of the simulations and also averaged over 100 runs, at three different parameter settings.

In the case of Fig. 7a the parameters ($d = 4$ and $TCA : x$) were set such that the end state of the social network showed polarisation, marked by two large communities. This meant that each agent was communicating with roughly 50 other agents, resulting in a large variety of topics being discussed. Therefore, the resulting distribution for the association strengths between the concepts lacks any large peaks and is not far from being homogeneous. In contrast, when the communities are small and the agents communicate only with a limited number of other agents, the shape of the association strength distribution drastically changes, as indicated by Fig. 7b. Here the parameters ($d = 3$ and $TCA : x^4$) were set such that the social network fragmented into a large number of smaller, densely connected communities. Due to the limited number of communication partners, under this setting a large peak forms in the association strength distribution around zero. This represents the topics (concept pairs) that come up only with minimal probability during communication. This formation is analogous to the emergence of "interest groups" in real social systems, which are smaller communities centered around specific or specialised topics. As we can see in Fig. 4c, this also gives slightly higher attitude homogeneity values inside the communities.

In order to gain an intuition about the shape of the distribution under extreme conditions we have also run simulations involving only a single pair of agents communicating with each other. Fig. 7c shows the association strength distribution of their belief systems at the end state. As it can be seen, reducing the number of communication partners to an extreme leads to a more uneven distribution, with association strengths either exceeding 0.6 or nearing 0. In this case, the number of topics which were involved in the communication was even more limited. A more detailed analysis of the two agent communication case can be found in the Supplementary Material Section S4.

Beside the shape of the association strength distribution, we have examined the similarity between the internal structure of the belief systems between different agents as well, with a special focus on agents sharing the same community. Along this line, we introduce the *belief system homogeneity*, *BH*, a quantity based on the comparison of the association strengths $B_{\alpha\beta}$ between the agents. For detailed definition see the Methods section. The belief system homogeneity calculated for a group of agents takes a value of $BH = 1$ if and only if all the association strengths are identical among the examined agents (e.g., like in the initial state consisting of identical agents), while lower values indicate diversity among the connection strengths in the different belief systems.

In Fig. 8 we show the *BH* at the end of the simulations among all the agents (blue) together with the average *BH* obtained for communities (red) as a function of the dissonance penalty *d* for the different *TCA* functions. We note that the value of $BH \simeq 0.7$, corresponding to the *BH* of the entire system at small and large dissonance

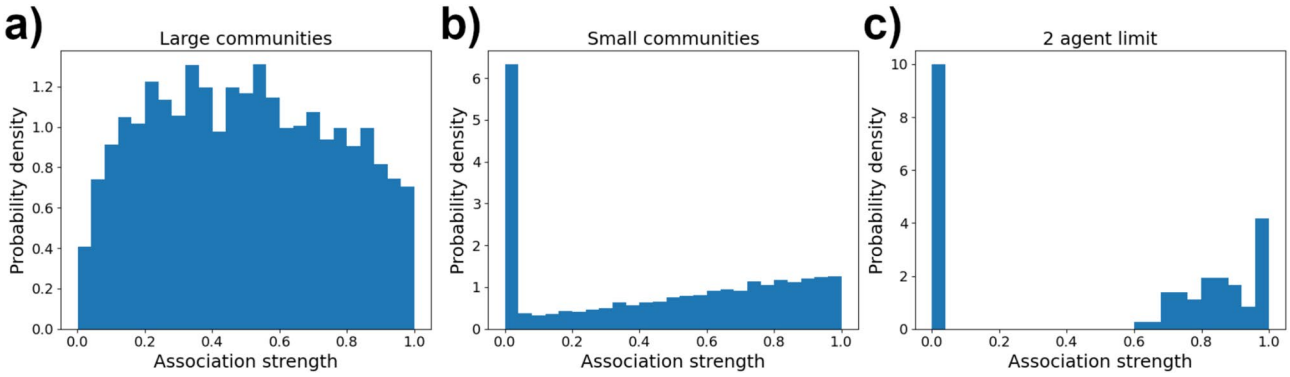


Fig. 7. Distribution of the association strengths in the belief system of the agents. Panel **a)** refer to the results for parameters $d = 4$ and $TCA : f(x) = x$ where the end state of the social network is composed of two large communities. Panel **b)** displays the association strength distribution of the belief systems at $d = 4$ and $TCA : f(x) = x^4$, where the communities typically have only 3 to 6 members, and are completely or almost isolated from the rest of the network. Finally, panel **c)** presents the results for special simulations where only a single pair of agents were communicating at $d = 4$ and $TCA : f(x) = x^4$. In the case of two agents and small communities, we can see that a large number of links disappear from the belief system. This means that the amount of topics agents will discuss during communication becomes smaller as well.

penalty values (i.e, $d < 2$ or $6 < d$) is slightly higher than the average *BH* in a null model where the $B_{\alpha,\beta}$ values are independently drawn uniform random variables in the$[0, 1]$ interval. This means that on average, the belief systems were slightly more similar to each other than if they were selected by random chance. According to Fig. 8, the *BH* inside the communities becomes larger compared to the *BH* in the entire system roughly between $d = 2.5$ and $d = 5$, corresponding to the range where the presence of communities in the social network is most significant according to the results described earlier. This effect is most prominent for the simulations using $TCA : f(x) = x^4$ (Fig. 8c), where a strong increase in the *BH* of communities is accompanied by a relevant drop in the *BH* of the entire system. This shows that the agents in the small communities that appear at this parameter setting develop belief systems that are very similar to the belief system of fellow community members, and in the meantime differ more from the belief system of agents outside the community. For the emergence of this scenario, see the supplementary video.

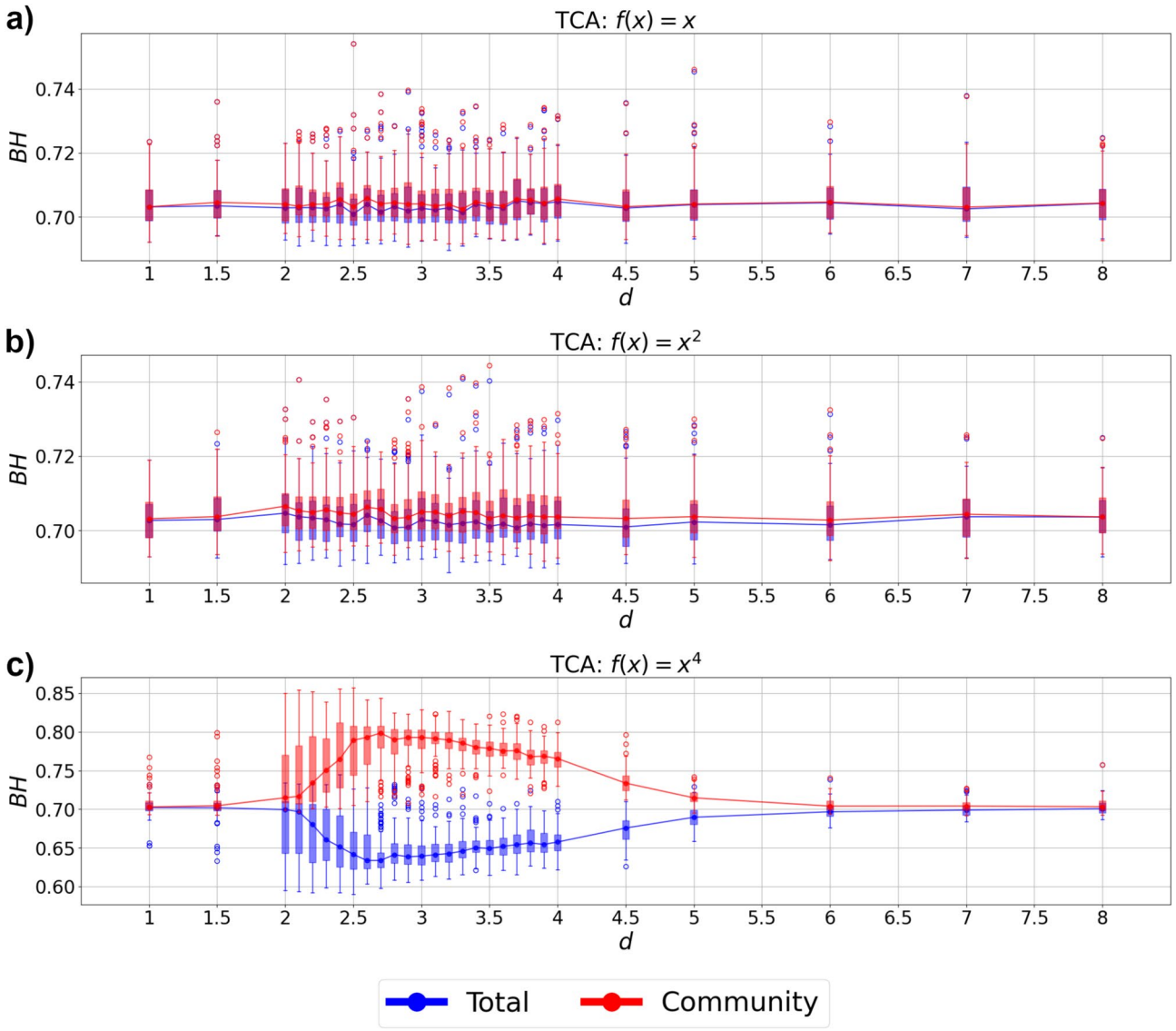


Fig. 8. Belief system homogeneity, *BH* , as function of the dissonance penalty, *d* . We show the *BH* calculated based on all the agents in the end state of the simulations in blue, whereas the average *BH* of communities is shown in red. The symbols represent a standard box plot over 100 runs for each parameter setting according to *d*, and the median values are connected by continuous curves. The applied *TCA* function is indicated in the panel titles. In panel **c)**, we can see a separation between the total *BH* and the community *BH* values. This coincides with the breakup of the social network in Fig. 6, when small separated communities begin to form, within which the belief systems of the members become more similar compared to that of agents in other communities.

Discussion

In the present paper we have proposed a social network model driven mainly by cognitive dissonance avoidance, corresponding to the keen effort for maintaining a coherent belief system. In order to study the effects this driving force can have on the structure of the social network, we have used an agent based approach where the agents communicate with each other along the social connections, and are also endowed with realistic human features, most prominently, by a belief system consisting of concepts (beliefs) that are associated with one another in an additional internal network. Similar models for the belief systems were already introduced by research communities interested in opinion dynamics, most prominently, but not exclusively[41], by those aiming to simulate the dynamics of political attitudes[38–40,83,84]. In a typical model for the internal network , the nodes represent "an element of a person's belief system"[40], which are named variously in different models, such as "opinions"[38], "concepts"[41] "attitudes"[38–40], "beliefs"[40,41,84], or "positions"[39,83]. The links in the network of the belief system correspond to the ties among the concepts, which in the traditional models are usually either positive or negative, depending on their supporting (consistent) or disproving (opposing) nature.

In contrast with earlier approaches, in the present work we assumed that the links between the attitudes can only be non-negative.

In parallel, we introduced node values in the internal network representation of the belief system for representing the attitudes of the agents towards the different concepts, which can be both positive or negative. Accordingly, the association between two concepts can either increase the coherency level of the belief system (providing the pleasant feeling of reassurance when the attitudes towards strongly connected concepts have the same sign), or inversely, it can create cognitive dissonance, which is decreasing the coherence of the belief system.

We analyzed the behavior of the proposed model through simulations that began with identical agents forming a fully connected social network. As the simulations progressed, both the connections within the social network and the internal belief systems of the agents underwent significant structural changes. The model incorporated two key parameters: the cognitive dissonance penalty, d, which represents the strength with which agents strive to avoid cognitive dissonance, and the *TCA* function, which governs the agents' tendency to initiate communication with the friends of their friends. The simulations were concluded once the social network's structure stabilized.

According to our results, in this end state of the social network – depending on the parameter setting – a wide range of social structures can appear, ranging from consensus through fragmentation to polarization. Consensus, where the social network remains dense, with all the agents forming a single group, is probably the most simple outcome we can expect based on the extremely homogeneous initial state of the simulations. This end state occurs at low values of d, where agents can tolerate more the possible cognitive dissonance that may emerge due to the communications between each other. However, at moderate d values and moderately increasing *TCA* functions (that is $f(x) = x$ or $f(x) = x^2$), where agents communicate with the friends of their friends with a relatively high probability, the social network becomes polarised, where a small number of large communities form. Members of these communities tend to have more similar belief systems among themselves compared to 'outsiders' both according to the attitudes and the structure of the association weights between the concepts. Last but not least, for a rapidly increasing *TCA* function of $f(x) = x^4$, where agents communicate almost only with their direct friends, we observed fragmentation of the social network in the moderate d regime, where a large number of small communities formed (composed of only a few agents each), that often became isolated from the rest of the network. This can be seen by the decrease in the size of the largest connected component. These small, isolated communities show a larger homogeneity in terms of the belief system of the members compared to the large communities mentioned previously.

The fact that a drastic change occurred in the structure of the social network roughly around $d \simeq 2$ for all studied *TCA* functions indicates that the cognitive dissonance avoidance built into our model acts as a very important driving force with a very strong impact on the emerging network structure. Furthermore, from the point of view of the belief systems, our results illustrate that this driving force can also create a wide range of often antagonistic and extreme attitudes even in initially completely homogeneous systems. Naturally, cognitive dissonance avoidance by itself would not induce the above effects for isolated agents. Instead the agents also have to be part of a social network in order to produce these interesting transitions, since it is the communication among the agents that sets any transition process afloat, by changing the strength of associations among the concepts (– which is a very basic representation of the mechanism of associative learning).

Furthermore, our finding that agents converge toward either uniformly positive or negative attitudes in the absence of fixed concepts suggests that internal coherence alone does not produce diverse or realistic belief landscapes. Instead, universal psychological anchors – such as fear of death or value of care – may be crucial for maintaining heterogeneity and resilience in belief systems. These results align with the predictions of existential psychology and support the inclusion of fixed-value concepts in theoretical modeling of belief dynamics.

While our model incorporates several interacting components, its central purpose remains conceptually focused: to explore how the effort to avoid cognitive dissonance – or equivalently, to maintain a coherent internal belief system – shapes both individual attitudes and the structure of social networks. This coherence-seeking drive is well established in psychology as a fundamental feature of human cognition, and our model is designed to trace its consequences in a structured and dynamic setting. The additional assumptions, such as triadic closure affinity or fixed positive/negative nodes, are meant to reflect empirically observed regularities in social interaction and moral valuation. However, the observed outcomes – from consensus to fragmentation – ultimately emerge from the tension between individual coherence and social exposure.

At the same time, it is important to note that cognitive dissonance can be decreased by various other ways as well (which methods are mastered by humans), for example by focusing on information that reassures the already existing ones while dismissing those that would create "unpleasant" associations – known as *confirmation*

bias –, or by accepting beliefs that "explain away" unpleasant associations (For example, somebody we love did something bad because she/he was forced to). This latter behaviour explains why intelligent people are so prone to accept strongly questionably information so easily, in case it fits well to their belief system. As a possible future direction, the inclusion of such cognitive dissonance decreasing strategies can also be incorporated into the model.

Another interesting extension is to allow variations among the agents in terms of which are the concepts that have a fixed constant positive or a fixed constant negative attitude in the belief system (representing communities with different value systems). Furthermore, the study of possible ways by which polarized communities can be depolarized, using the same framework can also lead to valuable insights.

Methods

Metrics

To help monitor the behaviour of our system, we defined several custom metrics. Their detailed explanation can be found below.

Extremism

To measure the extremism, *EM* of the attitudes of the agents, we can the sum of the absolute attitude values towards each belief, normalized by the belief system size *M*, written as

$$EM_i = \frac{1}{M} \sum_{\alpha} \left| a_{\alpha}^{(i)} \right| \tag{11}$$

The higher the extremism, the closer the attitudes are to the two poles of -1 and 1.

Attitude homogeneity

This metric is based on pairwise comparisons of the agents. For every pair, we take the sum of the absolute difference between the non-constant attitudes towards the same concepts. Then to normalize it, we divide by two times the number of non-constant attitudes, M^*, so that it has a maximum value of 1 if the two agents have completely opposite opinions, and 0 if they share the same ones. To transform this into a homogeneity metric, we subtract this value from 1, so that 1 means that the two agents have completely homogeneous beliefs. According to that, for a given pair of agents *i* and *j* we can write

$$AH_{ij} = 1 - \frac{1}{2M^*} \sum_{\alpha} \left| a_{\alpha}^{(i)} - a_{\alpha}^{(j)} \right|. \tag{12}$$

For any group of agents, such as a community or the entire social network, the attitude homogeneity, *AH*, is given by the average of AH_{ij} over all the possible agent pairs in the group.

Belief system homogeneity

To measure the similarity between the structure of the belief systems between the agents, we created a metric we call belief system homogeneity, *BH*, which is similar in nature to the attitude homogeneity. For a pair of agents, here take the sum of the absolute difference between each link weight in the belief system, then divide by the total number of links, given by $M(M-1)/2$. Following that we subtract the obtained value from 1, written as

$$BH_{ij} = 1 - \frac{2}{M(M-1)} \sum_{\alpha<\beta} \left| B_{\alpha\beta}^{(i)} - B_{\alpha\beta}^{(j)} \right|. \tag{13}$$

This way, a final result of $BH_{ij} = 1$ shows identical belief system weights, whereas 0 indicates the opposite. For groups of agents, the belief system homogeneity *BH* is simply given by the average of the BH_{ij} values over all possible pairs of agents within the group.

Communities

To calculate the number of communities in our social network, we used the Leiden [76] algorithm optimized for modularity included in the igraph Python package.[85] The Leiden algorithm [76] (widely known as the improved version of the Louvain approach [86]) works by initially assigning all agents their own communities. The nodes are then moved locally to neighboring communities in a way that maximizes an objective function, in the present case, modularity. Next, these communities are refined, meaning that the nodes inside them are merged into another partition, with the constraint that the new communities must be locally well connected. This refined partition is then aggregated into a new network, on which the same local moving and refinement steps are performed. The process then repeats, starting out from this new aggregated network until no further increase in the objective function is possible.

Modularity

Modularity is a metric used to determine the quality of a given network partition. It's defined as the difference between the fraction of edges found in communities and the expected fraction of edges inside the communities.[80] For a weighted network, this can be written in the following form:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta_{c_i, c_j}. \tag{14}$$

Here, m is the total weight of edges in the network, A_{ij} is the adjacency matrix containing the link weights, γ is the resolution parameter, k_i is the sum of link weights adjacent to node i, k_j is the sum of link weights adjacent to node j, δ is the Kronecker symbol, c_i and c_j are the partition of node i and j respectively.

t-SNE

T-distributed stochastic neighbour embedding (t-SNE) is a non-linear dimensionality reduction method.[81] The first step of the t-SNE algorithm is to calculate the probability that point i is the neighbour of point j in the high-dimensional space. This probability is defined as:

$$p_{j|i} = \frac{K(x_i, x_j)}{\sum_{k \neq i} K(x_i, x_j)}, \tag{15}$$

where, K is a Gaussian kernel function. A similar probability is also defined in the lower-dimensional space. Here instead of a Gaussian kernel, t-SNE uses the Student-t distribution, with $\nu = 1$:

$$q_{j|i} = \frac{(1 + |y_i - y_j|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + |y_k - y_l|^2)^{-1}}. \tag{16}$$

This probability is then optimized to be as close to the high-dimensional one as possible by minimizing the Kullback-Leibler divergence on the y values:

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{17}$$

The final result is a low-dimensional embedding, where the distances between the points are similar to those in the high-dimensional space.

UMAP

Uniform manifold approximation and projection (UMAP) is a non-linear dimensionality reduction method.[82] The algorithm approximates the manifold on which the high-dimensional data lies, and attempts to project it into a lower dimension. First, UMAP calculates the distances between the data points in the original dimension. Based on these distances, a radius is created around each data point, so that each of the radii includes at least the n nearest neighbours of the data point. This defines a cover, which is then made fuzzy by replacing the binary neighbour values with the following probability:

$$p_{ij}(X) = \exp\left(- \frac{d(x_i, x_j) - \rho_i}{\sigma_i} \right). \tag{18}$$

Here, X are the coordinates of the original data points, d is the distance between the two points, ρ_i is the distance to the nearest neighbour and σ is a normalizing factor, specific to i. This probability guarantees that each point is connected to at least it's first nearest neighbour. UMAP defines the low-dimensional connection probability as:

$$q_{ij}(Y) = \frac{1}{1 + a||y_i - y_j||_2^{2b}}. \tag{19}$$

Here, Y are the coordinates for every point in the low-dimensional space while a and b are hyperparameters. UMAP incorporates both an attractive and a repulsive force into its objective function, which ensures that similar points are closer to each other, while different points are further apart. The cross-entropy to be minimized for Y is defined as:

$$CE(X, Y) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} + (1 - p_{ij}) \log \frac{1 - p_{ij}}{1 - q_{ij}}. \tag{20}$$

Simulations

In our simulations, we first created the social network, with the given N and M values, as well as the list of constant positive and negative nodes in the belief system and the temperature constant. The further options included a flag for deciding whether the social network should initialize one random Agent and create N clones, or create all agents randomly. The value/range of the used parameters are listed in Table 2. After initialisation, we repeatedly simulate the communication between the agents, stopping the process once the following convergence criteria are met:

1. The sum of link weights in the social network shows no significant changes,
2. the subgroup belief homogeneity shows no significant changes.

Parameter	Values used during simulation
N - social network size	100 - 250
M - belief system size	10
T - stabilization temperature	0.01
b - number of stabilization iterations	10
Number of constant attitudes	2 (1 positive, 1 negative)

Table 2. Parameter values used during simulations. The value / range of the model parameters used in our simulation studies.

N	M	TCA	c_{edges}	$c_{beliefs}$	c_{count}
100	10	$f(x) = x$	$d/5000 < 0.5\%$	$d/p < 5\%$	200
100	10	$f(x) = x^2$	$d/5000 < 0.25\%$	$d/p < 5\%$	200
100	10	$f(x) = x^4$	$d/1000 < 0.1\%$	$d/p < 5\%$	200
250	10	$f(x) = x$	$d/10000 < 0.5\%$	$d/p < 5\%$	1000
250	10	$f(x) = x^2$	$d/10000 < 0.5\%$	$d/p < 5\%$	1000
250	10	$f(x) = x^4$	$d/1000 < 0.5\%$	$d/p < 5\%$	1000

Table 3. Convergence criteria during the simulations. The columns c_{edges} and $c_{beliefs}$ provide the threshold values used as convergence criteria for the summed link weights and community belief homogeneity respectively. *d* is the absolute difference from the previous 1000 data points and *p* is the mean of the previous 1000 data points as a normalization factor. c_{count} is the number of consecutive convergence checks required to end the simulation.

In each case, we compared the average of these measures in the last 1000 data points, taken at every 100th communication to the previous 1000 data points. For the link weights, we used constant normalization factors instead of the previous results to calculate the percentage change for faster convergence. To stop the simulation, the convergence criteria had to be fulfilled for c_{count} consecutive checks. The details of the convergence criteria are given in Table 3. The simulation results for 250 agent runs can be found in the Supplementary Information in Figs. S4-S7. These runs provided similar results to those discussed here in the main text.

Data availability

All data generated or analysed during this study are included in this published article.

Code availability

The code used for the simulations is available at https://github.com/ztoli17/Assoc-based-bsdyn.

Received: 9 October 2024; Accepted: 4 December 2025
Published online: 11 December 2025

References

1. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. https://doi.org/10.1103/RevModPhys.74.47 (2002).
2. Mendes, J. F. F. & Dorogovtsev, S. N. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, Oxford, 2003).
3. Newman, M. E. J. et al. (eds) *The Structure and Dynamics of Networks* (Princeton University Press, Princeton and Oxford, 2006).
4. Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125. https://doi.org/10.1016/j.physrep.2012.03.001 (2012) (**Temporal Networks**).
5. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical processes on complex networks* (Cambridge University Press, Cambridge, 2008).
6. Christakis, N. A. & Fowler, J. H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives* (Little, Brown, 2009).
7. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415 (2001).
8. DeGroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121 (1974).
9. Axelrod, R. *The Evolution of Cooperation* (Basic Books, 1984).
10. Schelling, T. C. *Micromotives and Macrobehavior* (W. W. Norton & Company; Revised edition, 2006).
11. Galam, S. *Sociophysics, A Physicist's Modeling of Psycho-political Phenomena* (Springer, 2012).
12. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).
13. Peralta, A. F., Kertész, J. & Iñiguez, G. Opinion dynamics in social networks: From models to data (2022). arXiv:2201.01322.
14. Perra, N. & Rocha, L. E. C. Modelling opinion dynamics in the age of algorithmic personalisation. *Sci. Rep.* https://doi.org/10.1038/s41598-019-43830-2 (2019).

15. Sîrbu, A., Loreto, V., Servedio, V. D. P. & Tria, F. *Opinion Dynamics: Models, Extensions and External Effects, 363–401* (Springer International Publishing, Cham, 2017).
16. Lorenz, J. Continuous opinion dynamics under bounded confidence: A survey. *Int. J. Mod. Phys. C* **18**, 1819–1838. https://doi.org/10.1142/S0129183107011789 (2007).
17. Galam, S. Sociophysics: A Review Of Galam Models. *Int. J. Mod. Phys. C (IJMPC)* **19**, 409–440. https://doi.org/10.1142/S0129183108012297 (2008).
18. Tóth, B. & Csányi, V. *Our Beliefs - The building blocks of human thoughts (in Hungarian: Hiedelmeink - Az emberi gondolatok építőkövei)* (Libri, Budapest, 2017).
19. Galef, J. *The Scout Mindset: Why Some People See Things Clearly and Others Don't* (Portfolio, 2021).
20. Sloman, S. & Fernbach, P. *The Knowledge Illusion: Why We Never Think Alone* (Riverhead books, New York, NY, 2017).
21. Shermer, M. *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies - How We Construct Beliefs and Reinforce Them as Truths* (St. Martin's Griffin; Illustrated edition, 2012).
22. Galotti, K. M. *Cognitive Psychology In and Out of the Laboratory* 5th edn. (SAGE Publications, 2013).
23. Dawkins, R. *The Selfish Gene* (Oxford University Press, 1976).
24. Lynch, A. *Thought Contagion* (Basic Books, 1998).
25. Converse, P. E. The nature of belief systems in mass publics. *Critical Review* **18** (1964).
26. Rokeach, M. The organization and modification of beliefs. *The Centen. Rev.* **7**, 375–395 (1963).
27. Grabisch, M. & Rusinowska, A. A survey on nonstrategic models of opinion dynamics. *Games* **11**, https://doi.org/10.3390/g11040065 (2020).
28. Galam, S., Gefen, Y. & Shapir, Y. Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *J. Math. Sociol.* **9**, 1–13 (1982).
29. Stone, M. The opinion pool. *The Annals of Mathematical Statistics* 1339–1342 (1961).
30. Chatterjee, S. & Seneta, E. Towards consensus: Some convergence theorems on repeated averaging. *J. Appl. Probab.* **14**, 89–97 (1977).
31. Häggström, O. & Hirscher, T. Further results on consensus formation in the deffuant model. *Electron. J. Probab.* **19**, 26. https://doi.org/10.1214/EJP.v19-3116 (2014).
32. Deffuant, G., Neau, D., Amblard, F. & Weisbuch, G. Mixing beliefs among interacting agents. *Adv. Complex Syst.* **03**, 87–98. https://doi.org/10.1142/S0219525900000078 (2000).
33. Axelrod, R., Daymude, J. J. & Forrest, S. Preventing extreme polarization of political attitudes. *Proc. Natl. Acad. Sci.* **118**, e2102139118. https://doi.org/10.1073/pnas.2102139118 (2021).
34. Latané, B. The psychology of social impact. *Am. Psychol.* **36**, 343–356 (1981).
35. Axelrod, R. The dissemination of culture: A model with local convergence and global polarization. *J. Confl. Resolut.* **41**, 203–226 (1997).
36. Raducha, T. & Gubiec, T. Coevolving complex networks in the model of social interactions. *Phys. A: Stat. Mech. Appl.* **471**, 427–435. https://doi.org/10.1016/j.physa.2016.12.079 (2017).
37. Zafeiris, A. Opinion polarization in human communities can emerge as a natural consequence of beliefs being interrelated. *Entropy* **24**, 1320. https://doi.org/10.3390/e24091320 (2022).
38. DellaPosta, D. Pluralistic collapse: The "oil spill" model of mass opinion polarization. *Am. Sociol. Rev.* **85**, 507–536. https://doi.org/10.1177/0003122420922989 (2020).
39. Brandt, M. J., Sibley, C. G. & Osborne, D. What is central to political belief system networks?. *Pers. Soc. Psychol. Bull.* **45**, 1352–1364. https://doi.org/10.1177/0146167218824354 (2019).
40. Brandt, M. J. & Sleegers, W. W. A. Evaluating belief system networks as a theory of political belief system dynamics. *Pers. Soc. Psychol. Rev.* **25**, 159–185. https://doi.org/10.1177/1088868321993751 (2021).
41. Rodriguez, N., Bollen, J. & Ahn, Y.-Y. Collective dynamics of belief evolution under cognitive coherence and social conformity. *PlosOne* **11**, https://doi.org/10.1371/journal.pone.0165910 (2016).
42. Galesic, M., Olsson, H., Dalege, J., van der Does, T. & Stein, D. L. Integrating social and cognitive aspects of belief dynamics: Towards a unifying framework. *J. Soc. Comput.* **1**, 278–292. https://doi.org/10.23919/JSC.2020.0004 (2020).
43. Dalege, J., Galesic, M. & Olsson, H. Networks of beliefs: An integrative theory of individual- and social-level belief dynamics. *Psychol. Rev.* **131**, 000–000. https://doi.org/10.1037/rev0000494 (2024).
44. Klein, S. B. *Learning: Principles and Applications* (SAGE Publications, Inc; 8th edition, 2018).
45. De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M. & Storms, G. The "small world of words" english word association norms for over 12,000 cue words. *Behav. Res. Methods* **51,** 987–1006. https://doi.org/10.3758/s13428-018-1115-7 (2019).
46. Hu, Z., Luo, J., Zhang, C. & Li, W. A natural language process-based framework for automatic association word extraction. *IEEE Access* **8**, 1986–1997. https://doi.org/10.1109/ACCESS.2019.2962154 (2020).
47. Cooper, J. *Cognitive Dissonance: 50 Years of a Classic Theory* (SAGE Publications Ltd, 2007).
48. Leonard, J. Cognitive dissonance: What to know. https://www.medicalnewstoday.com/articles/326738 (2019).
49. Bail, C. A. et al. Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci.* **115**, 9216–9221. https://doi.org/10.1073/pnas.1804840115 (2018).
50. Mäs, M. & Flache, A. Differentiation without distancing. explaining bi-polarization of opinions without negative influence. *PLoS ONE* **8**, e74516. https://doi.org/10.1371/journal.pone.0074516 (2013).
51. Hegselmann, R. & Krause, U. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artifical Societies and Social Simulation (JASSS)* **5** (2002).
52. Hebb, D. O. *The Organization of Behavior: A Neuropsychological Theory* (Wiley, New York, 1949).
53. Rescorla, R. A. & Wagner, A. R. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In Black, A. H. & Prokasy, W. F. (eds.) *Classical Conditioning II: Current Research and Theory*, 64–99 (Appleton-Century-Crofts, New York, 1972).
54. McClelland, J. L. & Rumelhart, D. E. Distributed memory and the representation of general and specific information. *J. Exp. Psychol.: Gen.* **114**, 159–188. https://doi.org/10.1037/0096-3445.114.2.159 (1985).
55. De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M. & Storms, G. The small world of words: English word association norms for over 12,000 cue words. *Behav. Res. Methods* **51**, 987–1006. https://doi.org/10.3758/s13428-018-1115-7 (2019).
56. Small World of Words Consortium. Small world of words project. https://smallworldofwords.org/en/project (2025). **Accessed: 6 May 2025**.
57. Goldstein, E. B. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience* (Cengage Learning; 5th edition, 2018).
58. Heider, F. Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946).
59. Estrada, E. Rethinking structural balance in signed social networks. *Discrete Appl. Math.* **268**, 70–90. https://doi.org/10.1016/j.dam.2019.04.019 (2019).
60. Talaga Szymon, T. J. S. . A. S. T., Massimo Stella. Polarization and multiscale structural balance in signed networks. *Commun. Phys.* **6**, https://doi.org/10.1038/s42005-023-01467-8 (2023).
61. Estrada, E. & Benzi, M. Walk-based measure of balance in signed networks: Detecting lack of balance in social networks. *Phys. Rev. E* **90**, 042802. https://doi.org/10.1103/PhysRevE.90.042802 (2014).
62. Bowlby, J. *Attachment and Loss, Vol. 1: Attachment* (Basic Books, 1969).
63. LeVine, R. A., Miller, P. M. & West, M. M. (eds.) *Parental Behavior in Diverse Societies* (Jossey-Bass, 1988).

64. Royle, N. J., Smiseth, P. T. & Kölliker, M. *The Evolution of Parental Care* (Oxford University Press, 2012).
65. Bourke, J. *Fear: A Cultural History* (Virago, 2006).
66. Becker, E. *The Denial of Death* (Free Press, 1997).
67. Hofstede, G. *Culture's Consequences* (Sage Publications, 1984).
68. Becker, E. *The Denial of Death* (Free Press, New York, 1973).
69. Solomon, S., Greenberg, J. & Pyszczynski, T. *The Worm at the Core: On the Role of Death in Life* (Random House, New York, 2015).
70. Haidt, J. *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (Pantheon Books, New York, 2012).
71. Rapoport, A. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *Bull. Math. Biophys.* **15**, 523–533 (1953).
72. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. https://doi.org/10.1038/30918 (1998).
73. Kossinets, G. & Watts, D. J. Empirical analysis of an evolving social network. *Science* **311**, 88–90 (2006).
74. Harmon-Jones, E. (ed.) *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology* 2nd edn. (American Psychological Association, 2019).
75. Easley, D. & Kleinberg, J. *Networks, Crowds, and Markets: Reasoning About A Highly Connected World* (Cambridge University Press, 2010).
76. Traag, V., Waltman, L. & van Eck, N. J. From louvain to leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233. https://doi.org/10.1038/s41598-019-41695-z (2019).
77. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002 (2010).
78. Fortunato, S. & Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **659**, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002 (2016) (**Community detection in networks: A user guide**).
79. Cherifi, H., Palla, G., Szymanski, B. & Lu, X. On community structure in complex networks: challenges and opportunities. *Appl. Netw. Sci.* **4**, 117. https://doi.org/10.1007/s41109-019-0238-9 (2019).
80. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113. https://doi.org/10.1103/PhysRevE.69.026113 (2004).
81. van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
82. McInnes, L., Healy, J., Saul, N. & Großberger, L. Umap: Uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861. https://doi.org/10.21105/joss.00861 (2018).
83. Baldassarri, D. & Goldberg, A. Neither ideologues nor agnostics: Alternative voters' belief system in an age of partisan politics. *Am. J. Sociol.* **120**, 45–95. https://doi.org/10.1086/676042 (2014).
84. Turner-Zwinkels, F. M., Johnson, B. B., Sibley, C. G. & Brandt, M. J. Conservatives' moral foundations are more densely connected than liberals' moral foundations. *Pers. Soc. Psychol. Bull.* **47**, 167–184. https://doi.org/10.1177/0146167220916070 (2020).
85. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Systems* **1695**(5), 1–9 (2006).
86. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.: Theory Exp.* **2008**, P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008 (2008).

Author contributions

A. Z. and G. P. developed the concept of the study, Z. K. implemented the model, carried out the simulations and performed the analyses. Z. K. prepared the figures, G. P. and A. Z. wrote the paper. All authors reviewed the manuscript.

Funding

This research was partially supported by the National Research, Development and Innovation Office – NKFIH, grant No. K128780 and by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-31643-5.

Correspondence and requests for materials should be addressed to G.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025