Universidad Zaragoza

Final Degree Project

EMERGENCE OF POLARIZATION IN COMPLEX NETWORKS

Author:
Alba María Téllez Fernández

Directors:
Dr. Jesús Gómez Gardeñes
Sergio Faci Lázaro

University of Zaragoza
Faculty of Sciences.
Department of Condensed Matter Physics.

Bachelor Degree in Physics.
Course 2018-2022

Contents

1 Introduction

Many academic problems studied in Physics focus on small-scale, linear, and Gaussian systems for specific pedagogical reasons. Mainly, because there exists well-known simple analytical and algorithmic solutions to linear systems. Gaussian statistics also allows working with the Central Limit Theorem and results in well-behaved functions. Moreover, finding a solution is more manageable with a small number of elements [1]. Unfortunately, nearly all major real-world systems do not fulfil some of these characteristics, as it is evidenced in most large-scale problems in biology, economics[2], epidemics [3], sociology [4, 5] or synchronization [6]. Physics of Complex Systems aims to address the difficulties that a detailed analysis of these multidisciplinary challenges present.

Numerous systems of scientific interest are composed of large numbers of interconnected dynamical units or *components* [7]. These interacting elements can range from humans and animals in societies and ecosystems to neurons in the brain. In all these systems, each component's functionality is affected by its connections with the rest of the system. Hence, we identify *complex systems* as large aggregates whose specific connection structure significantly influences the system's behaviour [1]. For instance, in social systems, we encounter thousands of interlinked human individuals as elementary constituents of the network. In a society, the *pattern of connections* influences how people learn, spread diseases, or create opinions.

The dependence between topology and functionality leads to essential properties observed in real-world networks from different fields. Despite their inherent differences, they share some unifying attributes. Some of the most relevant features a system may present to be termed complex are [8]:

- *Non-linearity*: Multiple systems must be described using non-linear interaction models. This property makes them subject to irreversible and discontinuous or catastrophic state changes.
- *Non-Gaussian*: Many systems can exhibit extreme behaviour that seems to be unpredictable based on past data [1]. Gaussian statistics are not appropriate to describe them as it assigns a zero-chance to radical occurrences where the underlying behaviour is indeed a power-law.
- *Non-regular interaction pattern*: The number of connections per system's element may differ significantly. This implies the presence of *modules* or community structures at a mesoscopic scale that lead to an enhancement of local synchronization in systems with no global order.

Depending on the system under consideration, these qualities may occur individually or collectively. Nonetheless, some individual properties of microscopic elements may get lost when analysing them as part of a complete system. For this reason, extrapolating features of a few elements does not always (if anytime) provide a faithful or accurate description of the entire system [9].

For each added level of complexity, new properties may emerge as a result of microscopic interactions. This way, the individual behaviour impacts the macroscopic behaviour [1]. P. W. Anderson made one of the first attempts to illustrate the need for new laws, concepts, and generalisations in the article "*More is different*" [9]. Anderson also describes how the interactions of the system's components can lead to unforeseen properties known as *emergent phenomena*. Physics of Complex Systems tries to explain how these collective behaviours may arise from the nonlinear interaction

of several small-scale components. Moreover, the Physics at hand is able to describe the system's topology and the nature of interactions without losing sight of the system's complex hierarchy [8].

In human societies, a collective effect of the interaction between individuals is the exhibition of astounding regularities on a large scale. People can establish a shared language or culture or even agree on controversial issues. The statistical physics approach has shown to be a very effective framework for describing this macroscopic social behaviour known as *global consensus*[10]. However, empirical observations do not always back up this prediction. Sometimes moderated initial opinions may evolve into a wide range of different points of view deviating from a global consensus [11].

Intuitively, we could explain the empirical data through the wide range of available media options that expose people to multiple viewpoints. On the one hand, this could allow people to understand a given topic from different perspectives and thus help them reach a consensus. On the other hand, members of society can be biased in favour of their own opinions and hence "pay more attention" to ideas they already agree with. Whenever individuals' interactions reflect a strong enough partisanship, *echo chambers* may emerge [11]. This phenomenon is characterised by individuals only interacting with those whose opinions are similar enough, reassuring their own beliefs. This biased selection of interactions in turn would entangle individuals in a concrete mindset [12], this could explain how a moderate stance can become extreme.

Lately, echo chambers and opinion polarisation have been quantified in several sociopolitical scenarios and through controversial debates on multiple social media platforms [13, 14, 15]. The motivation for studying these behaviours is their alleged relation to the spread of misinformation or falsehoods as well as their influence in public opinion. However, echo chambers contribute to a widening gap in information among the people interested in specific topics and those who are not [16]. In society, this manifests as segregation based on interests. Thus, everyone's opinion coincides or resonates with the one in their social circle and extreme opinions may show up.

In this Final Degree Project we not only address the classical question of dynamics reaching consensus but also opinion plurality in social systems. We will propose different mathematical approaches to provide insight into the transition between global consensus and the emergence of polarized states and echo chambers. Empirical evidence points to both *social influence* and *controversy* of the subject matter as the main elements that promote this transition, both off and online arguments [17]. It is therefore critical to use these two elements as a basis for describing the arise of polarization and echo chambers. With this purpose in mind, we will start by characterizing social systems as complex networks where social phenomena emerge from the pattern of interactions between humans. Our focus is on creating simple models that can reproduce complex macroscopic social outcomes from reasonably easy rules of behaviour in the fine-scale [18].

2 Complex networks

We can use complex systems to represent the pattern of interactions in our system of study, human societies, as well as in technological and biological systems [1]. The goal of this section is to show how complex networks models are constructed and offer mathematical tools to study their

dynamics. According to Graph Theory [19], the simplest approach to describe the global properties and topology of complex systems is to model them as a graph $\mathcal{G} \equiv (\mathcal{V}, \mathcal{L})$. This network is defined by a set of N nodes (vertices) $\mathcal{V} \equiv \{1, ..., N\}$ connected through some edges (links) $\mathcal{L}$. A pair of nodes i and j can only interact with each other if there is a link connecting them $(i, j) \equiv l_{i,j}$ [7].

In order to store the information of connected nodes, it is often useful to consider a matricial representation of a graph $\mathcal{G}(\mathcal{V}, \mathcal{L})$. This representation is known as the adjacency or connectivity matrix, $\mathcal{A}$, and it is a $N \times N$ square matrix whose entries A_{ij} $(i, j = 1, ..., N)$ represent the links of the network l_{ij}. If a node i has a link with node j, $i \longrightarrow j$, the element A_{ij} of the matrix would have a value different from 0 $(A_{ij} \neq 0)$. Otherwise, if the link l_{ij} does not exists, $i \not\longrightarrow j$, the element reads $A_{ij} = 0$. The diagonal of the adjacency matrix (whose elements are given by A_{ii}, for $i = 1, ..., N$) contains only zeros if there are no connections between a node i and itself, $i \not\longrightarrow i$. This way, depending on the values of the non zero elements, the adjacency matrix (and thus the network) may be classified as:

– **Unweighted or weighted network**

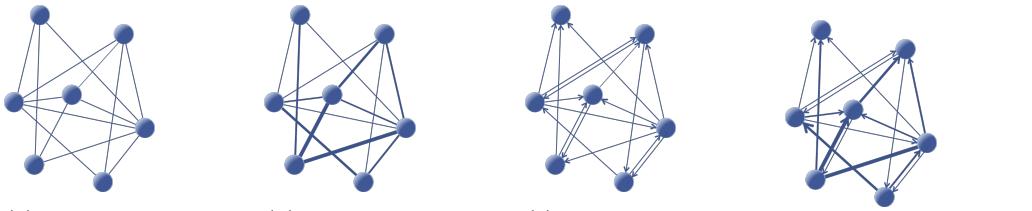
- Unweighted or Binary: These networks only take into consideration the existence or non-existence of links between nodes [7]. If there is a link between the nodes i and j, the element (i, j) of the adjacency matrix reads $A_{ij} = 1$. Otherwise, if the link does not exists, $A_{ij} = 0$.

- Weighted: Each link is defined by a pair of nodes i and j and a number, $w_{ij} \in \mathcal{R}$, called *weight* of the link. It represents the intensity of the connection inspired on the heterogeneous real-world interactions [7]. Here the not null elements of the adjacency matrix read $A_{ij} = w_{ij} \in \mathcal{R}$.

– **Directed or undirected network**

- Undirected: Each link is defined by a pair of nodes i and j where a connection $i \longrightarrow j$ implies also a connection $j \longrightarrow i$. The links l_{ij} and l_{ji} can be referred indistinctly. Therefore, the adjacency matrix is symmetric as its elements fulfil $A_{ij} = A_{ji}$ $\forall i, j \in N$ [7].

- Directed: Each directed link from node i to j is denoted by the ordered pair $l_{ij} = (i, j)$. Since the order of the nodes is relevant, the existence of a link l_{ij} does not necessarily imply l_{ji}, so $\mathcal{A}$ is not symmetric [7]. If a node i has an outgoing link to node j, $i \longrightarrow j$, the element (i, j) of the matrix reads $A_{ij} \neq 0$. Otherwise, if the link does not exists, the element reads $A_{ij} = 0$.



(a) Non-directed unweighted (b) Non-directed weighted (c) Directed unweighted (d) Directed weighted

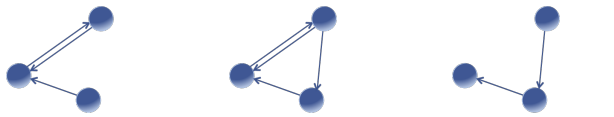
Figure 1: Graphical representation of non-directed unweighted (a) and weighted (b); directed unweighted (c) and weighted (d) graphs ($N = 7$ and $K = 14$ links). The arrows indicates the direction of each link and the weights are represented by the links' thicknesses.

– **Time non-dependant or dependant networks**

- Static or time non-dependant networks: Both the set of nodes and the links connecting them remain stationary over time. The matrix of adjacency is time independent $A_{ij} \neq A_{ij}(t)$

- Time-dependant: Graphs whose topological characteristics evolve with time by the addition and deletion of nodes and links or by changing the weight of the connections (i.e. the entries of the adjacency matrix $A_{ij} \equiv A_{ij}(t)$ change over time [7]). The Baràbasi-Albert model [20] (explained in Appendix A) is an example of a time-dependent network based on nodes changes. Whereas networks in which the evolution of connections is related to the dynamics of nodes, are an example of a time-dependent network based on links.

  $\Longrightarrow$ Adaptive Networks: These are a type of time-dependant networks. The topology changes on the network co-evolves with the dynamics governing the system under study, resulting in a feedback loop between the node dynamics and structure [21]. Such interplay is observed in a variety of circumstances, such as social or neural interactions. Adaptive networks are the basis for the study of social systems, and more specifically of this Final Degree Project.



(a) t (b) t+δ_t (c) t+$2\delta_t$

Figure 2: Adaptive network. Example of link evolution of the links through time

The networks will be encoded in the adjacency matrix, which can be studied to obtain more information and measure the networks' features.

2.1 Structural descriptors for complex networks

With the network encoded in the adjacency matrix, we can compute several metrics to learn more about the network's features. The simplest way to characterise and eventually distinguish the nodes of a graph is to count their number of links, i.e. to evaluate their *degree*. The degree (or *connectivity*) k_i of a node i is the number of direct unweighted connections from the node i to other nodes $j \in N$ of the network [22]. This number is defined in terms of the adjacency matrix as:

$$k_i = \sum_{j=1}^{N} A_{ij} \quad \text{with} \quad A_{ij} = 0 \text{ or } 1. \tag{1}$$

If the graph is directed, the degree of the node has two components: the number of outgoing links $k_i^{out} = \sum_{j=1}^{N} A_{ij}$ and incoming links $k_i^{in} = \sum_{j=1}^{N} A_{ji}$ so that the total degree is then: $k_i = k_i^{out} + k_i^{in}$. If the graph is weighted we can define the *strength* of a node i as the the sum of the weights of the edges connected to the node:

$$s_i = \sum_{j=1}^{N} A_{ij} \quad \text{with} \quad A_{ij} \in \mathcal{R}. \tag{2}$$

Also, we can define the parameter $\langle k \rangle$ as the *average connectivity* that quantifies the average number of neighbours of each node. Using Eq.[1] it can be represented as: [1]

$$\langle k \rangle = \frac{1}{N} \sum_{j=1}^{N} k_i = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij}. \tag{3}$$

Once the degree of each node in the network is computed, we can construct the *degree distribution* $P(k)$ defined as the probability that a node i chosen at random has degree k or, equivalently, as the fraction of nodes in the graph having degree k [22].

The degree distribution provides a description of the complexity of the whole graph and, as we show below, it can be radically different depending on the network model at work. Depending on the type of underlying degree distribution we have, networks can be classified into different groups:

- *k-Regular networks*: All nodes have the same connectivity value of k. In other words, every node has the same number k of neighbors.
- *Random networks:* These networks are characterized by the disordered nature of the links' arrangement between different nodes. Since all nodes in a random network are statistically equivalent, the probability that a link is established between two given nodes is constant ($p \in [0, 1]$) and the same for all the $N(N-1)/2$ pairs. For large N and a particular value of p the average connectivity of a node is $\langle k \rangle = p(N-1)$, and the degree distribution is well approximated by a Poisson distribution [22]: $P(k) = e^{\langle k \rangle} \cdot \langle k \rangle^k / k!$.
- *Scale Free networks*: Although Poissonian degree distributions or other types of homogeneous functions for $P(k)$ are recurrently found in random network models, real networks show a rather different behavior and their degree distributions appear to be heterogeneous and display fat tails. Therefore, $P(k)$ significantly deviates from the Poisson distribution and, in many cases, exhibits a power law (scale-free) distribution: $P(k) \propto k^{-\gamma}$ with $\gamma \in [2, 3]$ This power law distribution allows a small fraction of nodes (*hubs*) to be connected to an elevate number of nodes. Whereas, the rest of elements show a low-value connection degree.

The degree distribution can be used as a basis for the construction procedure for networks. Apart from the degree of a node and the degree distribution of a network, other important descriptors are *clustering coefficient* (probability that two nodes connected to the same third node are also connected to each other) and *shortest path length* (average distance between each pair of nodes in the network)[22]. In Appendix A we briefly discuss three of the most generic and simple models of networks: Erdös and Rényi [23], Watts-Strogatz [24] and Barabàsi-Albert [20].

3 Modeling echo chambers

Once introduced the basis for constructing and characterising a complex network, now use it to describe a social system and investigate the dynamics that governs its evolution. In order to represent the relationships on a social system, we will use an adaptive network where the changing pattern of connections is reflected both on the dynamical units behaviour and the adjacency matrix

[1] In directed graphs, Eq.[3] is $\langle k \rangle = \langle k_{out} \rangle = \langle k_{in} \rangle$

[25]. We will show how the use of competitive adaptive mechanisms gives a hint of the emergence of the echo chamber phenomena and the polarization in society.

3.1 Social systems as complex networks

A social network is essentially defined as a group of people or social entities interconnected with each other through some type of interaction. Friendships, partnerships, sexual encounters, and commercial ties are some examples of interactions. In these networks, the vertices portrait the people and are named "agents or actors"; whereas the social interactions between them are represented by the edges and are known as "ties" [1].

In Section 2 we explained the different types of networks so now, we will shift our focus towards the one that best suits the problem of modeling social systems addressed in this project: directed adaptive networks. In our first approach to model social systems we will use binary networks and, after that, we will shift to weighted networks. Hence, we will now go through the criteria on which these networks' structural patterns co-evolve with the dynamics of the nodes, so the connectivity pattern of the agents is not static but dynamic. This mechanism of co-evolution is determined by the competition of two adaptive principles:

(i) Reinforcement of those interactions within synchronized units. This principle is known in sociology as ***homophily*** [26] and it plays a key role in the spreading of information and the emergence of culture in social networks.

(ii) Each unit has at its disposal a finite amount of the available resources (e.g. time) to interact with the rest of the network. This restricts the number of links that an unit can have simultaneously, limiting its associative capacity. The principle that rules this mechanism is known as ***homeostasis*** [27] and it is related to the Dunbar's number [28].

For modeling social networks we will consider a system of N interconnected agents. Each element is described by a one-dimensional variable representing its opinion about a certain topic at each time step, $x_i(t) \in (-\infty, +\infty)$. The first approach for modeling the opinions' evolution over time was proposed in the article [16] by F. Baumann *et al.* which tries to capture the social interaction dynamics as we will explain in Section 3.2. As the system will be characterized by the variable opinion and its evolution, we will first start by giving the physical interpretation of the sign and module of the variable $|x_i|$.

The sign of x_i, $\sigma(x_i)$, denotes agent i's ***qualitative attitude*** towards a binary decision in which there are only two alternatives (p,e: in favour or against). Positive signs $\sigma(x_i) > 0$ are related to affirmative of positive opinions (p,e: in favour) and $\sigma(x_i) < 0$ to negqtives ones (p,e: against). When the opinion's value of an agent i is null $x_i = 0$, and thus, $\sigma(x_i) = 0$, it means that this agent i has not opted for any of the alternatives. The opinion $x = 0$ is called the ***neutral opinion*** or state. Another useful measure is the absolute value of the opinion x_i of agent i, $|x_i|$. This value measures the ***conviction strength*** of i towards dissenting opinions from other nodes. The i-th agents attitude towards its position becomes increasingly extreme as the module $|x_i|$ grows larger.

3.2 Opinon's Dynamics

Following [16] we assume that the opinion's dynamics are purely driven by agent interactions, and model it using a system of N coupled ordinary differential equations for each of the agents:

$$\dot{x_i} = -x_i + \lambda \sum_{j=1}^{N} A_{ji}(t) \cdot \tanh(\alpha \cdot x_j), \tag{4}$$

where $\lambda > 0$ defines the *social interaction strength* within agents and $A_{ji}(t)$ is the temporal adjacency matrix whose evolution will be later discussed in Section 3.3.

The *influence function* is introduced in Eq. 4. This function characterises people's capacity to influence other individuals' opinions towards their own. It is taken to be a sigmoid function controlled by the parameter $\alpha > 0$: $\tanh(\alpha x)$. The particular nonlinear shape of these impact function complies with the prior qualities evidenced by empirical results [17] as it ensures that:

(*i*) An agent j is able to sway other's in the direction of its own opinion's sign $\sigma(x_j)$, since $\tanh(x)$ is an odd function [2] (See Fig. 3).

(*ii*) The social influence that one agent i has over others peers grows monotonously with the conviction $|x_j|$ of the agent, as $|\tanh(\alpha \cdot x_j)| \sim |x_j|$.

(*iii*) Analytically, the parameter $\alpha > 0$ controls the degree of non-linearity between an agent's opinion and the social influence they have on other individuals.
Since the influence function is a sigmoid function, the influence of extreme opinions has a limitation (as it eventually saturates when $\tanh(\alpha x) \approx 1$); and α controls the range of opinions x_i that are considered non-extreme. Whenever one opinion is consider radical, its impact on other nodes gets saturated regardless of its conviction $|x_j|$.
Additionally, α could be interpreted as the controversy of the topic at hand. Larger values of α are related to more contentious topics, in which fewer perspectives are labeled as non-extreme. More explicitly:

- Small values of α suggest that moderate individuals exert a little social influence on their peers. The moderate opinion range is huge because it is a lighthearted subject. This way, opinions not differing much from the neutral opinion will have a minimal impact.
- Large values of α indicate that even moderately opinionated agents can have a substantial social effect on others. If there is a lot of controversy around the topic, the moderate opinion range is smaller since things are highly sensitive with this issue. In the limit of $\alpha \longrightarrow \infty$, no matter how little the opinion differs from what the mainstream thinks, this position will be regarded as radical and its influence will be limited.

It is noteworthy the fact that we represent opinion dynamics as a fully collective, self-organizing process with no inherent individual preferences. As a result, the opinions of agents with no social contacts decrease toward the neutral state.

[2] Odd function: $\tanh(-x) = -\tanh(x)$

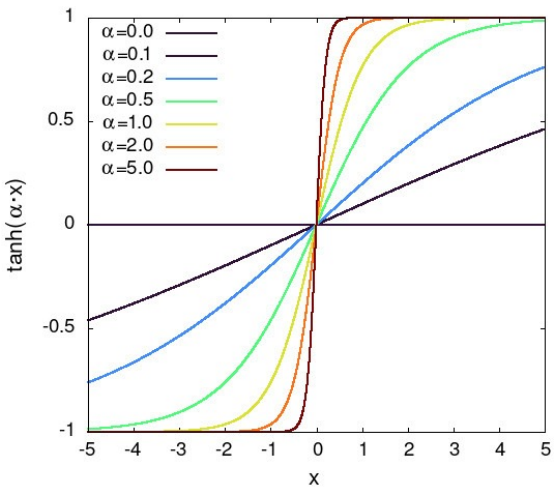


Figure 3: Here we represent the influence function with $\alpha > 0$. If the function was null, there would be no influence from others at all. The influence of the opinions reaches it maximum when $\tanh(\alpha x) = 1$ for a value of an opinion threshold x_θ. If another opinion x is $x > x_\theta$, the influence does not increase (saturation). As α increases, x_θ is closer to 0 and saturates with less extremist opinions. In the limit of $\alpha \longrightarrow \infty$, the function $\tanh(\alpha x) \longrightarrow \sigma(x)$ approaches the sign function, -1 for $x < 0$ and $+1$ for $x > 0$. This corresponds to a society with maximal social influence.

3.3 Activity-driven (AD) temporal network

With the dynamics of agents' opinions described in the previous Section 3.2, we now focus on the evolution of the networks' interactions. As we already discussed, the link structure of an adaptive network is encoded in a time-evolving adjacency matrix $A_{ji}(t)$. In a social network, the element of the matrix $A_{ji}(t)$ determines the i-th agent's neighbours, with a value of $A_{ji}(t) = 1$ if there is a connection from node j to i at time t. Otherwise, if there is not a connection, $A_{ji}(t) = 0$.

In general, information flow on social media is uneven, with a degree of asymmetry varying depending on the social media platform in question. Here, we analyse a network with directed interactions[3] where the directed interaction between i and j $(i \to j)$ has a probability r of being reciprocal: $Prob(j \to i | i \to j) = r$. If agent i receives social feedback from j: $A_{ij}(t) = A_{ji}(t) = 1$.

Following empirical observations, an activity driven model [29] is proposed to describe the instantaneous adjacency matrix. F. Baumann *et al.*[16] define an *activity rate* that represents the likelihood or propensity that agent i gets involved in m distinct random contacts with other agents at each time step. All the activity potentials of each node are obtained from an activity distribution $F(a)$, which, as suggested by empirical data [30], is usually assumed to follow a power law:

$$F(a) = \frac{1-\gamma}{1-\epsilon^{1-\gamma}} \cdot a^{-\gamma} \propto a^{-\gamma} \text{ with } a \in [\epsilon, 1]. \tag{5}$$

This distribution is defined as the probability that a randomly chosen agent i has an activity potential a. To avoid the scenario of having agents that are never active, we impose a lower cut-off, ϵ, for a. This way, we avoid any divergences of $F(a)$ near the origin.

The basis of the AD dynamics are completely determined by the parameters (ϵ, γ, m). The numerical implementation of the algorithm is detailed as follows:

(1) The network starts with N disconnected vertices and at each discrete time step t.

(2) At each time step, the i-th agent becomes active with a probability a_i.

[3]If an agent i establishes a link to another agent j, $i \to j$ does not necessarily imply a reciprocal link from $j \to i$.

(*3*) An active agent i influences m distinct agents $\{j\}$. This influence is expressed by a directed link $(i \to j)$ in the temporal adjacency matrix, *i.e.* $A_{ij}(t_n) = 1$. This step reflects homeostasis (see Section 3.1), since there is a limit m in the number of links that an agent can have simultaneously. An agent i chooses those m new peers j, either active or non-active, according to connection probability p_{ij}. This probability is modeled as a decreasing function of the distance between their opinions[4]:

$$p_{ij} = \frac{|x_i - x_j|^{-\beta}}{\sum_j |x_i - x_j|^{-\beta}} \tag{6}$$

In Eq.[6], the exponent $\beta > 0$ controls the power-law decay of the probability p_{ij} with opinion distance $|x_i - x_j|$. This introduces homophilic effects (see Section 3.1) in the new agents' interactions and therefore β is called the *homophily exponent*. The more similar j's view is to i's, the more likely it is to be chosen by i. Furthermore, β indicates the importance of the reinforcement of interactions within similar opinions. For larger values of β, the probability of being chosen having similar views (small opinion distance) increases, consequently increasing its reinforcement. In addition, this model also introduces the probability r that each link made by an active agent may be reciprocal.

(*4*) Opinions x_i are updated by numerically integrating Eq. 4 using $A_{ij}(t_n)$. The mathematical integration is done using an explicit fourth-order Runge-Kutta method and a time step of δt.

(*5*) At the next time step $t+\delta t$, all the edges in the network are deleted. This way, all interactions have a constant duration of δt.

3.4 Results

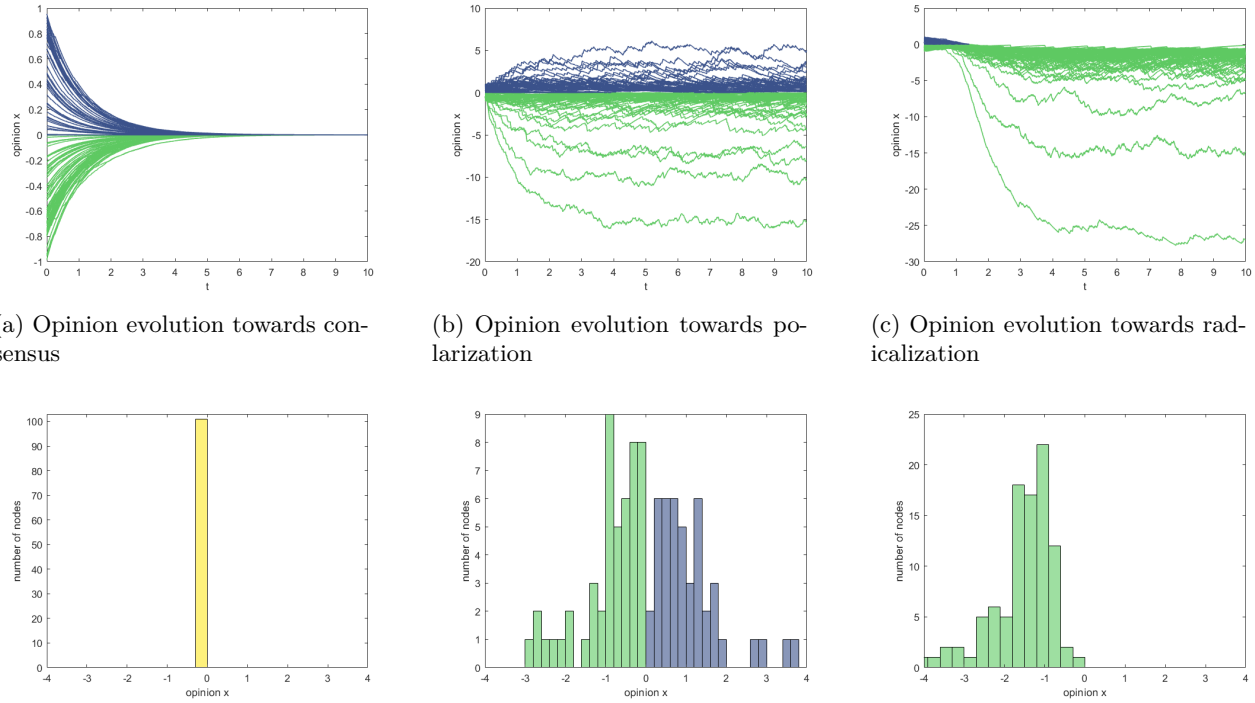
We analyze the model's behavior in terms of the social interaction strength K, how controversial is the issue at hand α, and the homophily exponent β. For each simulation, we initialize the opinions in the interval $x_i \in [-1, 1]$ according to a random uniform distribution. It can be demonstrated that the model does not critically depend on the initial conditions. The emergence of consensus and opinion polarization is recovered from initially balanced and highly asymmetric distributions of opinions [16].

Letting the parameter α be fixed, and varying the parameters K and β we are able to identify three qualitatively different dynamical regimes regarding the distribution of the final opinions x_f:

- Consensus: For low values of K, regardless of β, we see a convergence of x_f to the neutral opinion's value as the only surviving opinion (Fig.4(a)). This is shown in the opinion distribution as one remarkable peak on the neutral state $x_f = 0$ (Fig. 4(d)).
- Polarization: When both K and β are large we see how the initial opinions become more extreme far away from the neutral state, both with positive and negative sign (Fig. 4(b)). At the final state, the surviving opinions are essentially distributed around two peaks, one positive and another negative. This leads to a bimodal x_f distribution (Fig 4(e)).

[4]This is a change/departure from the uniform random selection that the original AD formulation proposed.

- Radicalization: A big value of K but with low β favors extremization of the initial opinions, far away from the neutral state either with positive or negative sign. For example, in (Fig. 4(c)), we observe a sharp decline of positive final opinions and also how the negative x_i's get radical values. We see how the surviving opinions have a value far away from the neutral state, resulting in a one-sided radicalized state. Similarly to consensus, radicalization has one peak on the final opinion distribution (Fig 4(f)). However, it is not a global convergence, as the peak is not sharp.



(a) Opinion evolution towards consensus

(b) Opinion evolution towards polarization

(c) Opinion evolution towards radicalization

(d) Final opinion: consensus

(e) Final opinion: polarization

(f) Final opinion: radicalization

Figure 4: Evolution and histograms of the different final states. Consensus appears for low K, regardless of β. High K gives rise to polarization and radicalization states, where a high β favors the former. Concretely the values implemented for each panel are: (a) consensus: $K = 0.5$, $\beta = 2$; (b) polarization: $K = 3$, $\beta = 3$; and (c) radicalization: $K = 3$, $\beta = 0$. Contentiousness and reciprocity are set to $\alpha = 3$ and $r = 0.5$, respectively, for all panels. Positive (negative) opinions are colored in blue (green). For more details regarding the simulation values see Appendix B

3.5 Shortcomings of the model

The proposed model gives a hint of the emergence of echo chamber phenomena and polarization in society. It is capable of replicating some empirical aspects seen on polarized social networks:

- Individuals in society interact differently and participate in heterogeneous activities. Users that are more active and involved in social interactions have a higher likelihood of displaying more extreme opinions than the rest [16].
- The opinions of a person and their interacting social network neighbours are frequently similar.
- When moderately held beliefs interact in a group, they might become extreme [31] as evidenced in conformity and imitation research [16].

- We can distinguish between consensus, radicalization and polarization in social opinion.

The ability to study collective phenomena in the population should characterise the dynamics of social networks. However, despite the achievements of this model to represent polarized states, we did not find it to be a faithful replica of reality for the following reasons:

(*i*) There is no evolution in the adjacency matrix. The adjacency matrix is generated at each instant of time based on the evolution of opinions. Unlike matrix evolutions, it does not evolve over time, but rather changes arbitrarily.

(*ii*) Since the AD model generates an adjacency matrix $\mathcal{A}$ based on the opinions at each time step, the dynamics of the matrix depend only on the state of the last time step. This means the AD model can be classified as Markovian model, and therefore, agents do not have memory of the previous time steps.

(*iii*) The lifetime of polarized states is finite [16], although it increases at least exponentially with the strength of homophily β, up to a point where the destabilization becomes numerically inaccessible.

In the following sections we will elaborate a new model that aims to solve the previously mentioned problems. To this aim, we will introduce a model of an adaptive network whose adjacency matrix $\mathcal{A}$ evolves with time rather than generating a new one in each time step. This type of time-evolving adaptive networks has mostly been studied in coupled oscillator models, based on the Kuramoto model (KM). Therefore, in the next section we will explain the KM model and, afterwards, we will see how to include memory in the evolution of the network. Finally, we will adapt this evolution of the matrix to the dynamics of the previously described opinions.

4 Time-evolving adaptive adjacency matrix

The study of social systems as complex networks aims to describe the collective behaviour in a population which arises from the interplay of the network pattern of connections and the opinion's distribution. Since the pattern of connections changes with time, we require an adaptive network to model the opinions' dynamics, as we already did in Section 3.3 with the AD temporal network. However, in the previous adaptive network proposed, the adjacency matrix that governs the opinions' dynamics is generated at each time step instead of evolving over time. The empirical data shows that the evolution of the opinions should be governed by the competition between the mechanisms of homophily and homeostasis (See Section 3.1). These mechanisms are thought to operate in a wide variety of real-world systems, and traditionally have been studied in synchronization of large populations of coupled phase-oscillators.

Most analytical studies of adaptive networks of oscillators that reflect homophily and homeostasis are based on the simple and widely studied nonlinear phase oscillator model: the Kuramoto Model (KM)[32]. Therefore, we will first explain the basis of the KM and its further improvements in order to implement memory and the mechanisms we are interested in. The analytical study of these oscillators' models will allow us to develop new tools and give us an insight about the

emergence of collective behaviour. We will later adapt the underlying interplay between network architecture and the distribution in the oscillators' natural frequencies to evolve the opinions in a social system.

4.1 Kuramoto Model

The Kuramoto model [32] corresponds to the analytical description of a system of N oscillators with the simplest possible case being the system with all-to-all interactions and equally weighted couplings. In these systems, we consider an ensemble of N limit-cycle oscillators with phase θ_i with $i = 1, .., N$ solely characterized by their natural frequency ω_i.

On a longer time scale, each unit exerts a phase-dependant influence on the other units, changing their rhythm, $\dot{\theta}_i$. Therefore, the dynamics of the system is given by the phase time evolution of the n-th oscillating unit that reads:

$$\dot{\theta}_n(t) = \omega_n + \lambda \sum_{m=1}^{N} \sin(\theta_m(t) - \theta_n(t)), \quad n = 1, ..., N \quad \text{with} \quad \lambda = K/N \tag{7}$$

where the interaction between oscillators is ruled by the sinusoidal coupling and a *constant coupling strength* λ. We include in λ the constant coupling weight $K \geq 0$ and $1/N$ to avoid divergences in the thermodynamic limit ($N \longrightarrow \infty$). The natural frequency of the n-oscillator, ω_i is given by a probability density $Prob(\omega_n \in (\omega, \omega + d\omega)) = g(\omega)$, usually assumed to be unimodal and symmetric about its mean frequency $\langle\omega\rangle \equiv \Omega$, like Lorentzian (Cauchy) and Gaussian distributions. Due to the rotational symmetry in the model, we can set $\Omega = 0$, without loss of generality, by redefining $\omega_n \longrightarrow \omega_n - \Omega \quad \forall\, n$. This corresponds to using a rotating frame at the mean frequency Ω [5] so that the ω_i's denote deviations from Ω. The governing equations Eq.[7] remain invariant but we had effectively shift the mean of $g(\omega)$ to zero [6].

Oscillators can display the dynamical analogue of a phase transition [33, 34]. As long as the spread of the natural frequencies is large compared to the coupling λ, each oscillator n moves at its own frequency θ_n and the system behaves incoherently. However, when a critical coupling threshold λ_c is crossed, collective synchronization appears. The oscillators freeze into a common mode despite their different individual frequencies. In order to measure the coherence of the system, we define the Kuramoto order parameter.

4.1.1 Kuramoto order parameter

The collective dynamics of the whole population ($n = 1, ..., N$) is quantified by the macroscopic quantity known as Kuramoto order parameter:

$$r(t) \cdot e^{i \cdot \phi(t)} = \frac{1}{N} \sum_{m=1}^{N} e^{i \cdot \theta_m(t)} \tag{8}$$

Analogously to a mean-field approximation in physics, this complex quantity $r(t) \cdot e^{i \cdot \phi(t)}$ can be interpreted as the collective rhythm produced by the whole population. The modulus $0 \leq r(t) \leq 1$

[5] $\theta_i \longrightarrow \theta_n + \Omega t$

[6] i.e.:$g(\Omega + \omega) = g(\Omega - \omega) \quad \forall\, \omega$ with $\Omega = \frac{1}{N} \sum_{m=1}^{N} \omega_m$. From now on $g(\omega) = g(-\omega) \ \forall \omega$.

measures the phase coherence and $\phi(t)$ is the average phase. The value of $r \approx 1$ states that exists a collective rhythm and $r \approx 0$ that there is incoherence, i.e: no macroscopic rhythm (See Fig 5).

4.1.2 Mean-field analysis

Kuramoto explored stable solutions using a mean-field analysis where the parameter $r(t)$ is constant, r, and the mean phase $\Phi(t)$ rotates uniformly at frequency Ω. The governing equations Eq.[7] can be rewritten in terms of the order parameter r (See Appendix C.1). This states that each oscillator interacts with all the others only through the mean field quantities r and ϕ. Furthermore, it exposes the positive feedback loop between each oscillator rhythm with the system's collective rhythm. The individual phases θ_n are pulled towards the mean phase Φ rather than toward the phase of any other individual oscillator.

The effective strength of the coupling is proportional to the coherence r, i.e., to the fraction of oscillators frozen in synchrony (positive-feedback loop). As long as the coupling between the oscillators is strengthened in each turn, more of them tend to be recruited into the synchronized set. The solutions of the mean-field governing equations show two types of long-term behaviour depending on how large are all the n-th frequency values $|\omega_n|$ in comparison to $\lambda \cdot N$:

- The group of oscillators fulfil that $|\omega_n| \leq \lambda N \cdot r^*$, they are phase-locked at frequency Ω and are approaching the stable fixed point: $\omega_i = \lambda N \cdot r^* \cdot sin(\theta_i^*)$, $|\theta_i| \leq \pi/2$ (Appendix C.1).
- The rest of oscillators, for which $|\omega_n| > \lambda N \cdot r^*$ is verified, are drifting in a non-uniform manner around the circle, sometimes accelerating and sometimes rotating at lower frequencies.

4.1.3 Transition to synchronization

The coupling value λ changes the fraction of phase-locked and drifting oscillators, which varies the collective rhythm given by the Kuramoto order parameter $r(t) \in [0, 1]$, Eq.[8]. To describe quantitatively how synchronized the system is expected to be in the long-term depending on its coupling λ, we use numerical simulations as the ones plotted in Fig.5(a)-5(d) and we will study the stationary value of the Kuramoto order parameter r^* .

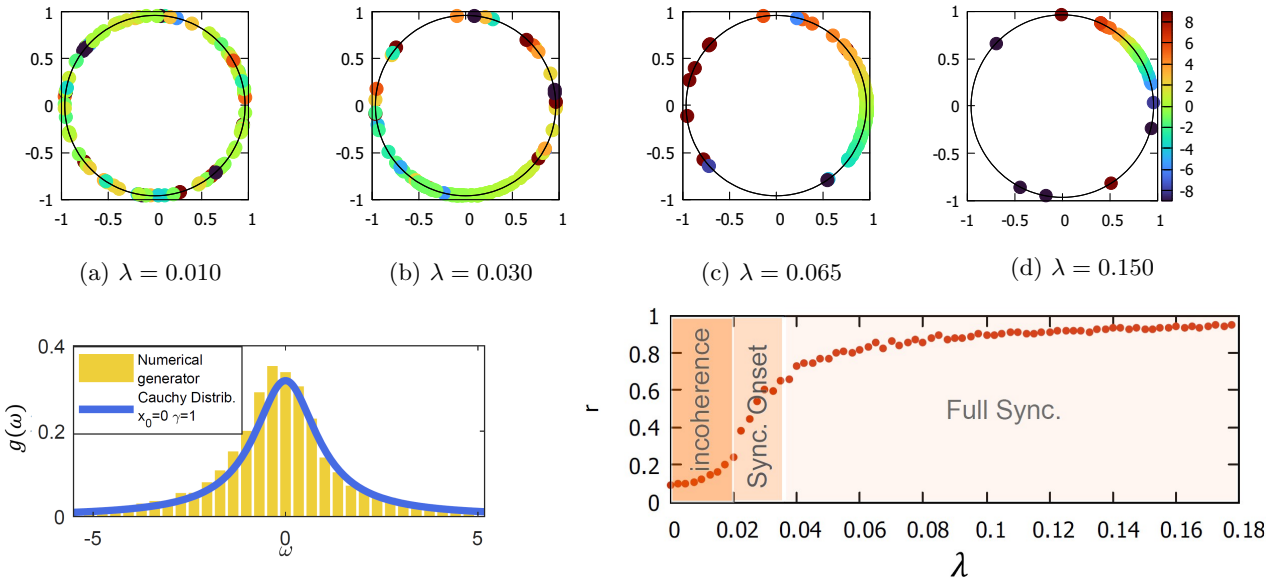
Starting from a value $\lambda \approx 0$ where almost no oscillator satisfy the phase-locked condition as $\lambda N \cdot r^* \approx 0$, oscillators appear to be uncoupled (drifting) and the phases are evenly spaced out around the unit circle, as we can see in Fig.5(a). In this incoherent state, r^* decays near to a small residual value of size $\mathcal{O}(1/\sqrt{N})$. If we slightly increase λ, despite $\lambda N \cdot r^* > 0$, as $r^* \approx 0$, the fraction of phased-locked oscillator is very small and the dynamics of each oscillator will be (practically) independent from the rest so that the system will remain incoherent. Eventually, the coupling λ will reach a critical value λ_c where the influence of each oscillator with the rest causes the system to start synchronizing as seen in Fig.5(b). Here, r^* is significantly larger than 0 (comparatively larger than the $\mathcal{O}(1/\sqrt{N})$ fluctuations). It reflects a partially synchronized state known as onset of synchronization where two groups in the population emerge:

- The oscillators near the center of the frequency distribution get mutually synchronized and get involved in a collective oscillation. These oscillators lock together at the mean frequency Ω and co-rotate with the average phase $\Phi(t)$.

- In the meanwhile, those oscillators in the tails of $g(\omega)$ run their natural frequencies and drift relative to the synchronized cluster. The more we increase λ, more oscillators get into the synchronized cluster and r^* grows regardless of the initial conditions.

No synchronization is attained below the critical coupling $\lambda < \lambda_c$ (sub-critical regime) and for $\lambda > \lambda_c$, the uncoupled state becomes unstable and r^* grows exponentially, increasing the fraction of phase-locked oscillators (Fig.5(c)) as nearly all the oscillators satisfies the locking condition $|\omega_n| << \lambda N \cdot r^*$ which causes r^* to increase at the same time. This process continues until the Kuramoto order parameter saturates at some value $r^* \approx 1$ (still with $\mathcal{O}(1/\sqrt{N})$ fluctuations) where the system is fully synchronized as in Fig.5(d).

We simulated in Fig.5(f) the value r^∞ at which the Kuramoto order parameter r^* saturates for a given λ from a fixed ω initial distribution $g(\omega)$ (Fig.5(e)). This offers the quantitative study of the phase transition into synchronization seen qualitatively on the numerical simulations plotted on Fig.5(a)-5(d):



(a) $\lambda = 0.010$ (b) $\lambda = 0.030$ (c) $\lambda = 0.065$ (d) $\lambda = 0.150$

(e) ω Cauchy distribution (f) Evolution of r

Figure 5: To visualize the dynamics of the system we represent each oscillator n as $e^{i\cdot\theta_n}(t)$. The phases $\theta_n(t)$ are shown as dots running around the unit circle in the complex plane and the centroid is given by the complex number $r \cdot e^{i\Phi}$ where r measures phase coherence and Φ is average phase. Each oscillator has an assigned color representing its natural frequency ω_n. In images Fig.5(a) to Fig.5(d) We can see how as the coupling λ increases, the oscillators start to stick together (synchronization) and therefore the coherence r increases. In Fig.5(b) $r \approx 0$ (individual oscillations add incoherence, there is no macroscopic rhythm) and in Fig.5(d) $r \approx 1$ (population with collective rhythm). Simulations for Kuramoto's model ($N \longrightarrow \infty$). The Lorentzian distribution $g(\omega)$ that gives the initial value of ω is shown in Fig.5(e). In Fig.5(f) we present the synchronization diagram, i.e., the Kuramoto order parameter r as a function of λ.

The transition from partially synchronized $r_\infty \in \mathcal{R}$ to entirely synchronized ($r_\infty \to 1$) seems to be a second order transition. The fully synchronized dynamics start to emerge after a certain large

enough value of the coupling ($\lambda >> \lambda_c$) and near λ_c, r_∞, obey the usual square-root scaling law for mean field models. The critical coupling is the minimum value of λ, from which synchronized dynamics emerge. In Appendix C.3 we show how the mean-field allows the calculation of λ_c:

$$\lambda_c = \frac{2}{N\pi \cdot g(0)} \tag{9}$$

Doing this with our simulation we obtain $\lambda_c \approx 0.02$ that is proven numerically in Fig.5(f). We should point that for studying the critical values of the KM we used an all-to-all and equally weighted coupling so that the adjacency matrix $\mathcal{A}$ of the network is static, non-weighted and non-directed. However, the connection topology of the real systems, like social systems, usually do not show a homogeneous pattern of connections as required for the mean-field approach. They usually exhibit a complicated topology of links that we encode as a complex network using an adjacency matrix $\mathcal{A}$ (See Section2).

4.1.4 Kuramoto Model in complex networks:

The KM was indeed one of the first dynamical processes to be studied with a interaction structure of a complex network. To accomplish so, it was necessary to take the links between oscillators into account in Eq.[7]:

$$\dot{\theta}_n(t) = \omega_i + \lambda \sum_{m=1}^{N} A_{nm} \cdot \sin(\theta_m(t) - \theta_n(t)), \quad n = 1, ..., N \tag{10}$$

where $A_{nm} > 0$ stands for the strength of the link from node j to i. The coupling strength between pairs of connected oscillators is assumed to be identical for simplicity. However, if the coupling value changes for every link ($n \to m$), then $\lambda \to \lambda_{nm}$. The original Kuramoto model (Eq.[7]) is recovered by letting $A_{nm} = 1 \ \forall n \neq m$ (all-to-all) and $\lambda_{nm} = \lambda \ \forall n, m$ (identical coupling strength).

Using the Time-Average Theory [35] we consider that every oscillator is influenced by the local field created in its neighborhood. For this reason, a local order parameter is defined as:

$$r_n \cdot e^{i\cdot\Phi} = \sum_{m=1}^{N} A_{nm} \cdot \langle e^{i\cdot\theta_m(t)} \rangle_t \tag{11}$$

where $\langle ... \rangle_t$ stands for a time average. Introducing r_n in Eq.[10] the steady state reads as: $\omega_n = \lambda \cdot r_n^* \cdot sin(\theta_n^*(t) - \phi_n)$ (See Appendix. C.2). Therefore, our locking condition reads $|\omega_i| \leq \lambda r_i$.

Using the mean-field approximation we define a new global order parameter $\tilde{r}$ to measure the global coherence. Moreover, we assume that the local parameter of every node n is proportional to its degree k_n (Section 2), so that $r_n = \tilde{r} \cdot k_n$. When the network is close to the onset on synchronization, $\lambda \longrightarrow \lambda_c^+$, $\tilde{r} \longrightarrow 0^+$. This leads to the condition for critical coupling (See Appendix C.4):

$$\lambda_c = \frac{2}{g(0) \cdot \pi} \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{12}$$

We simulated the case of $N = 100$ with a Cauchy distribution for the frequencies so that $\lambda_c \approx 0.0215$ which corresponds with the numerical simulation done in Fig.5(f). Below this critical value $\lambda < \lambda_c$, Eq.[12], no synchronization is attained and $r(t)$ decays to a small residual value of size $O(1/\sqrt{N})$ (sub-critical regime). While for $\lambda > \lambda_c$ the order parameter reaches a stationary value. The critical coupling λ_c in complex networks is the corresponding all-to-all topology λ_c (Eq.[9]) re-

scaled by the ration between the first two moments of the degree distribution. Note that it means that the collective behaviour transition strongly depends on the underlying network structure $\mathcal{A}$.

The dynamics of oscillators with complex pattern of connections is regulated by the KM, which is a paradigmatic framework for studying collective behaviour observed in real-systems, p.e., societies. Now that the link structure has been included in Eq.[7], we are going to include the mechanism that reshapes the network structure governed by the competition between homophily and homeostasis (see Section 3.1). This way, the weights of the adjacency matrix $\mathcal{A}$ will evolve with time, changing from time-independent to time-dependant $\mathcal{A} \equiv \mathcal{A}(t)$. Furthermore, we will introduce an adaptive weighted network that allows the system to retain memory.

4.2 Kuramoto Model with memory

In this section we are going to introduce a simple model of an adaptive network of phase oscillators in which the weights of the adjacency matrix $\mathcal{A}(t)$ co-evolve with the dynamics of the system. The matrix entries evolve with an adaptive scheme which has the form of the replicator equation of evolutionary dynamics [36] and retains the main characteristics of both homophily and homeostasis. These characteristics lead to the emergence of structural and dynamical features observed in real-systems such as social systems. For this reason, we present these adaptive scheme, introduced by S.Assenza *et al.* in the article [37], that will be later used to improve the opinions' dynamics Eq.[4] introduced in Section 3.2:

$$\dot{A}_{nm}(t) = A_{nm}\left[s_n \cdot p^T_{nm}(t) - \sum_{l=1}^{N} A_{nl}(t) \cdot p^T_{nl}(t)\right] \tag{13}$$

In Eq.[13], s_i stands for the total incoming strenght of node n, Eq.[2] and $A_{nl}(t)$ is the element (n, l) time-dependant adjacency matrix. The parameter $p^T_{nm}(t)$ represents the degree of local synchronization between oscillators i and j, averaged over time in the interval $[t - T, t]$. The possible synchronized links in the network in this time interval are computed as:

$$p^T_{nm}(t) = \left|\frac{1}{T}\int_{t-T}^{t} \exp\left(i\left[\theta_m(\tau) - \theta_n(\tau)\right]\right) \cdot d\tau\right| \tag{14}$$

In Eq.[14], T can be seen as a control parameter that quantifies the amount of memory used by each oscillator for taking into consideration the previous dynamics of the rest of the graph in its own updating process. The quantity $p^T_{nm}(t)$ gives us information about how locally synchronized the oscillators n and m have been along the last time units T. This variable is normalised by dividing it by T so it is bounded in the interval $[0, 1]$: the value $p^T_{nm}(t) = 1$ means that the input node m has been perfectly synchronized along all the last T time units with the target node n; meanwhile, $p^T_{nm}(t) = 0$ exposes the absolute lack of synchronization between the studied nodes.

More precisely, focusing on Eq. 13, we observe that the m inputs would be synchronized with the target node n if the value of its degree of coherence p_{nm} is strictly higher than the summation of all the degree of coherence between the nodes that n is connected to. In other terms, the m oscillators that are highly synchronized with the n-th oscillator in the last time interval fulfill:

$$p_{nm} > \sum_{l} A_{nl} \cdot p_{nl} \tag{15}$$

In addition, if we carefully analyse the inside of the brackets in Eq.[13] we are able to observe how the main characteristics of the self-regulating process carried out by homophily and homeostasis affect the synchronization of the oscillators:

- Those m inputs that in the last T time units have been highly synchronized with the target node n (they fulfil Eq.15) will increase the value of $\dot{A}_{nm}(t)$ in Eq.[13]. Therefore reinforcing those interactions with other correlated units in the graph as they enhance their strength. This corresponds to the so called homophily phenomena.
- The remaining nodes are not synchronised enough in the last T time units, they do not fulfill Eq.[15]. In accordance with the homeostasis phenomena, their weights will be decreased in order to preserve the total incoming strength value s_n of the oscillator n received by each unit, which initially was set to $s_n = 1\ \forall n$ in $t = 0$. Regarding Eq. 13, homeostasis arise when the strength does not evolve in time, $\dot{s_n} = \sum_{m=1}^{N} \dot{A}_{nm} = 0$. In practice this implies that all the links pointing to the same target i compete for the available resources.

This way, the system maintains its own stability while adapting its dynamics to varying conditions. In order to show that consistently, we quantify the system's global synchronization through the time-dependent Kuramoto order parameter (See Section 4.1 Eq.[8]). This way, we will be able to see how the system evolves towards an asymptotic state.

4.3 Results

The network starts with a random k-regular network of N oscillators each one with a fixed number K of connections. Thus, every element (n, m) of the adjacency matrix $\mathcal{A}$ has an initial weight of $A_{nm} = 1/K$. The evolution of the network reads as follows:

- For $t < 0$, we integrated numerically Eq.[7] the homogeneous network. Here the weights do not change in time and are fixed to $A_{nm} = 1/K$ for all links.
- For $t \geq 0$ we considered the full dynamics of the adaptive model by switching on the weights' evolution governed by Eq.[13]. When the weights co-evolve with the oscillators' dynamics, a clear enhancement of synchronization is observed for any values of λ and T.

We now explore in Figure 6 the time-evolution of the order parameter $r(t)$ with different values λ and fixed T. Prior to the introduction of the evolution of the adjacency matrix ($t < 0$) with Eq.[13], we had either incoherent or synchronized dynamics, depending on whether the value of the coupling is lower or greater than λ_c (see Section Eq.[9]). Once Eq.[13] (and Eq.[14]) is introduced, for couplings under the critical coupling $\lambda < \lambda_c$ we expect incoherent dynamics ($r(t) << 1$), Fig.6 $\lambda = 0.1$. After the onset of synchronization, for relatively large values of λ, Fig.6 $\lambda = 0.75$, we expect the the synchronization in going to highly increase, obtaining a perfectly full synchronized state ($r \approx 1$), obtaining values of $p_{nm} \longrightarrow 1$. This way, regarding Eq.[13], the weights of $\mathcal{A}$ will remain similar to its initial values $A_{nm} = 1/K$, and the resulting network is thus very similar to the initial one. However, as we increase the coupling λ before the critical coupling $\lambda < \lambda_c$, Fig.6 $\lambda = 0.5, 0.65$, we observe that a partially ordered phase emerges with a value of $\langle r(t) \rangle_t \approx 0.5$. In this phase, the network splits into into two clusters and is mainly separated in two components.

This mechanism has the unusual consequence of generally increasing synchronisation for strengths below the threshold limit.

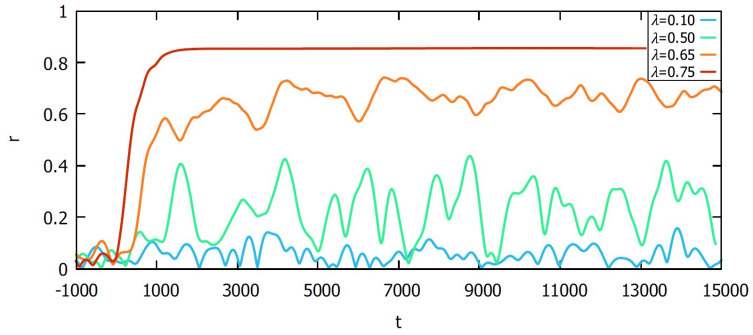


Figure 6: Time evolution of the global order parameter r for $\lambda = \{0.10, 0.50, 0.65, 0.75\}$ and memory $T = 100$. For $t < 0$ we can see the dynamics for a static network where the adjacency matrix does not evolve in time depending on the coupling λ. For $t > 0$, the weights evolve according to Eq.[13]. Also we notice quasi-periodic behaviour of $r(T)$ for the particularly ordered states in $\lambda = 0.5, 0.65$.

In the following final section of the project, we will examine how to incorporate the adaptive evolution of the adjacency matrix introduced in Section 4.2 into the model of opinion evolution examined in Section 3.2. In order to carry this out, we must make a number of adjustments in order to make them compatible while also trying to improve them.

5 Modeling echo chambers with adaptive networks and memory

In Section 3.2 we introduced a mathematical approach for modelling the dynamics of opinion purely driven by agent interactions. We discussed a system with N agents whose opinions x_i were characterised by their stance, $\sigma(x_i)$, and their strength, $|xi|$. In this model, when the agents interact with one another, their opinions evolve over time under the control of Eq. 4. This dynamic was ruled by the parameters λ and α. Where λ represents the node's social activity strength between agents and α the contentiousness of the subject matter on which the agents are expressing their opinions.

Systems where participants are more likely to change their minds and have controversial topics will be associated with higher values of λ and α, respectively. This hints that both, λ and α represent the same degree of freedom: the social influence of opinions on the agents. This made us hypothesize that the dynamics of the system would not significantly change if the values of these variables were increased or decreased, respectively. This is demonstrated by the fact that in F. Baumann *et al.* [16], systems with the same value of λ and α exhibit the same behaviour, even during phase transition.

For this reason, we believe that both parameters really reflect the same degree of freedom in the opinion's dynamics (the influence of one opinion on the agents). As a result, the governing equation would be simplified if we redefined opinions as $\tilde{x} = \alpha \cdot x$. By doing this, we obtain the new parameter:

$$\tilde{\lambda} = \lambda \cdot \alpha. \tag{16}$$

We refer to $\tilde{\lambda}$ as the effective social influence that each node j has over the connected node i. As a result, the system is slightly simplified since we now have one control parameter less. Substituting the opinion value x_i of the node i with $\tilde{x}_i = \alpha \cdot x_i \quad i = (1, ..., N)$, and introducing this change in Eq. 4 results in the new equation that rules the opinion dynamics is:

$$\dot{\tilde{x}}_i = \tilde{x}_i + \tilde{\lambda} \sum_{j=1}^{N} A_{ji}(t) \cdot \tanh(\tilde{x_j}) \tag{17}$$

In Section 3.3 we introduced the evolution of the adjacency matrix using an Activity Driven model. As we said before, this was not a realistic adaptive evolution. Thus, we adapt the evolution of the adjacency matrix of oscillators in the Kuramoto Model with memory T explained in Section 4.2: the network starts with a regular network of size N nodes with each one having K connections so that every element of the adjacency matrix has an initial value of $\frac{1}{K}$. At each time step t_p, the active agents (entries of the adjacency matrix A) change according to Eq.[13]:

$$\dot{A}_{ij}(t) = A_{ij} \left[s_n \cdot p_{ij}^T(t) - \sum_{l=1}^{N} A_{il}(t) \cdot p_{il}^T(t) \right]$$

Therefore, The agent n takes more or less into consideration the opinion of m according to how similar both of the opinions in the last T time is according to p_{nm}^T. However, instead of using the expression for p_{ij}^T given by Eq.[14], it is changed to:

$$p_{ij}^T = \frac{1}{T} \int_{t-T}^{t} e^{-\beta |x_i - x_j|} dt \tag{18}$$

Where the *homophily exponent* $\beta > 0$ controls the power-law decay of the probability p_{ij} with opinion distance. For larger values of β, the probability of being chosen having similar views (small opinion distance) increases and consequently increasing its reinforcement.

To obtain the opinions evolution, we will have to solve numerically the system of N equations that give the redefined opinions $\tilde{x_i}$ (Eq.17) with $i \in [1, N]$ and the $N \cdot N$ equations that gives the evolution of the adjacency matrix Eq.[13]. The mathematical integration is done using an explicit fourth-order Runge-Kutta method and a time step of δt. Now, the basis of the new dynamics is fully determined by the parameters: the effective social impact,$\tilde{\lambda}$, the homophily exponent, β and the memory, T, to keep track of others agent's opinions over this time. By changing these parameters we will study the final state of each resultant network using some network descriptors now introduced.

5.1 Descriptors for opinion polarization

In order to have comparable results throughout all the simulations we normalized the final opinions dividing every value of $\tilde{x_f}$ by the maximum value of $|\tilde{x_f}|$. This gives us normalized opinions defined by:

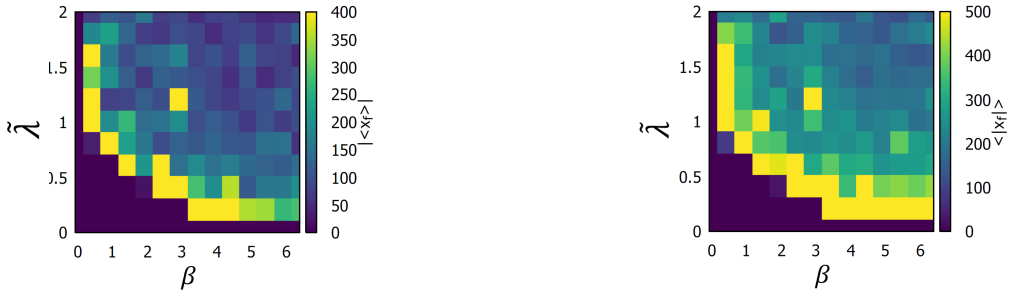
$$\hat{x}_{f,i} = \frac{\tilde{x}_{f,i}}{max(|\tilde{x}_{f,i}|)} \qquad i \in (1, ..., N) \tag{19}$$

For the maximum value of $|\tilde{x_{f,i}}|$ we did not consider opinions that deviated more than $5 - \sigma$'s deviation from the absolute value of the final opinions mean $|\langle \tilde{x_f} \rangle|$. This helps avoid highly extreme values deceiving our results.

The transition from neutral consensus to radicalization or polarization is depicted on Fig. 7 in the $(\beta, \tilde{\lambda})$-plane. The color encodes the absolute value of the final average opinion and the average of the absolute value of the final opinions:

$$|\langle \hat{x}_f \rangle| = |N^{-1} \sum_i \hat{x}_i(t_{final})| \tag{20}$$

¡In the long-term regime, the value of $|\langle \hat{x}_f \rangle|$ identifies the transition between regions exhibiting a stable neutral consensus and those where radicalization and polarization emerge and become stronger.



(a) Absolute values of the average final opinions $|\langle x_i \rangle|$

(b) Average absolute value of final opinions $\langle |x_i| \rangle$

Figure 7: Transition from consensus to extreme opinions dynamics. $|\langle x_i \rangle|$ and $\langle |x_i| \rangle$ in the $\beta - \tilde{\lambda}$ for $K = 10$ and $N = 100$ In the dark region, the system approaches a neutral consensus, while in the brighter areas the population undergoes dynamics where extreme opinions appears.

In Fig.7, neutral consensus regions, $|\langle \hat{x}_f \rangle| = \langle |\hat{x}_f| \rangle = 0$ (dark purple), are characterized by small β and $\tilde{\lambda}$ values. Increasing a small value (β, $\tilde{\lambda}$) we observe a clearly defined transition between consensus and highly extreme opinions dynamics. Here, both $|\langle \hat{x}_f \rangle|$ and $\langle |\hat{x}_f| \rangle \longrightarrow 1$ (yellow). We see how for a large homophily exponent β, the reinforcement of the self opinions is highly successful as the social influence from the rest of the population is low $\tilde{\lambda}$, leading to extreme opinions. Also, if the social influence is large $\tilde{\lambda}$ but the reinforcement of their own opinions, β, is low, the most extreme opinions will have a great influence on the population, leading to more extreme x_i. As we start increasing (β, $\tilde{\lambda}$), these values become gradually less extreme decreasing its values till $|\langle \hat{x}_f \rangle|$ and $\langle |\hat{x}_f| \rangle \rightarrow 0$ without reaching the consensus (green to blue). For instance, if we have a large homophily exponent β, the self-reinforcement is very relevant, but if the social influence $\tilde{\lambda}$ is also large so that "everyone listen to everyone", less extreme opinions will emerge.

However, we can not fully describe at what state is the system for every pair of values in the β-$\tilde{\lambda}$ plane. Therefore, in the following subsections we defined new variables in order to describe the systems' state with a better perspective.

5.1.1 The consensus fraction

An agent is considered to be in a consensus state if its final opinion is in the neutral state or close to it $|\langle \hat{x}_f \rangle| \in [0, \epsilon]$ with a small enough ϵ. We define the fraction of nodes in the network that are in a consensus state as ρ_0:

$$\rho_0 = \frac{\sum_{j=1}^{N} \Theta(\epsilon - |\hat{x}_f, j|)}{N} \tag{21}$$

Where $\Theta(x)$ is the Heaviside function [7] This measure clearly shows the transition from neutral consensus to radicalization or polarization in the $\beta, \tilde{\lambda}$-plane (See Fig. 8(a)). Neutral consensus regions have the maximal value of ρ_0 according to Eq. 21.The $\rho_0 = 1$ region (yellow), is characterized by small values of β and $\tilde{\lambda}$. In the regions where radicalization and polarization emerge and becomes stronger, almost all opinions fulfil $|\langle \hat{x}_f \rangle| > \epsilon$ (color coded purple), so that $\rho_0 \approx 0$ is obtained for increasing β and $\tilde{\lambda}$.

5.1.2 The majority fraction

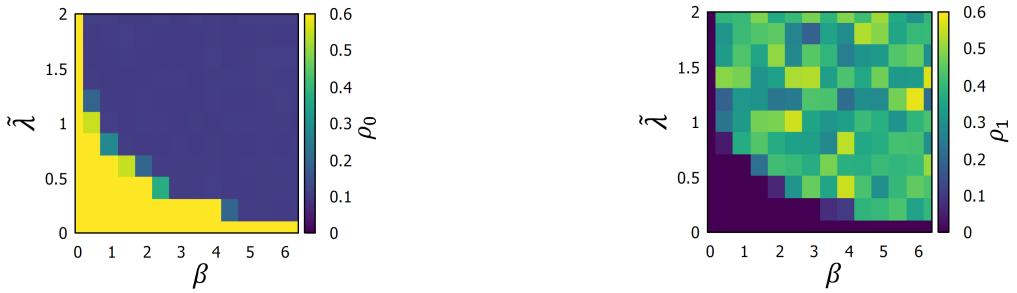
Once we have clearly identified the consensus state, we study the sign $\sigma(\hat{x}_f)$ of the agents' opinions. We study the opinion's fraction that are out of consensus with a positive (ρ_+) or negative (ρ_-) sign:

$$\rho_+ = \frac{\sum_{j=1}^{N} \Theta((\hat{x}_f, j) - \epsilon)}{N} \qquad \rho_- = \frac{\sum_{j=1}^{N} \Theta((\hat{x}_f, j) - (-\epsilon))}{N} \tag{22}$$

After calculating these measures, we define ρ_1 as the majority's fraction of numbers representing positive or negative opinions divided by the number of nodes out of consensus:

$$\rho_1 = 2 \cdot \left(\frac{max(\rho_+, \rho_-)}{\rho_+ + \rho_-} - \frac{1}{2} \right) \tag{23}$$

The parameter ρ_1 not only shows the transition from neutral consensus to extreme opinion dynamics in the $\beta, \tilde{\lambda}$-plane (See Fig. 8(b)), but also quantifies the degree of radicalization of the system: in a completely radicalized state this measure takes the value $\rho_1 = 1$; whereas in a completely symmetrical polarized state, it would take the value of $\rho_1 = 0$.



(a) ρ_0 (b) ρ_1

Figure 8: Neutral consensus regions have a value of $\rho_1 = 0$ according to 23 (dark blue). Whereas the $\rho_1 > 0$ region (green to yellow) indicates the non consensus region.

Nevertheless, in the Fig. 8(b) almost none of the cases are fully radicalized $\rho_1 \approx 1$ (yellow) or polarized $\rho_1 \approx 0$ (dark purple). We have plenty of values that have an intermediate value of ρ_1 (green). This means that only with ρ_1 we are not able to describe the agents' state when they are not in consensus, we cannot clearly say if they are radicalized or polarized. For this reason, we introduce another indicator for the radicalization.

[7] $\Theta(x) = 1$ si $x > 0$ o $\Theta(x) = 0$ si $x < 0$.

5.1.3 The radicalization descriptor

Once we know through ρ_1 the biggest fraction of non-consensus opinions, we define an indicator which represents how far from the consensus these opinions are. In other words, how extreme they are:

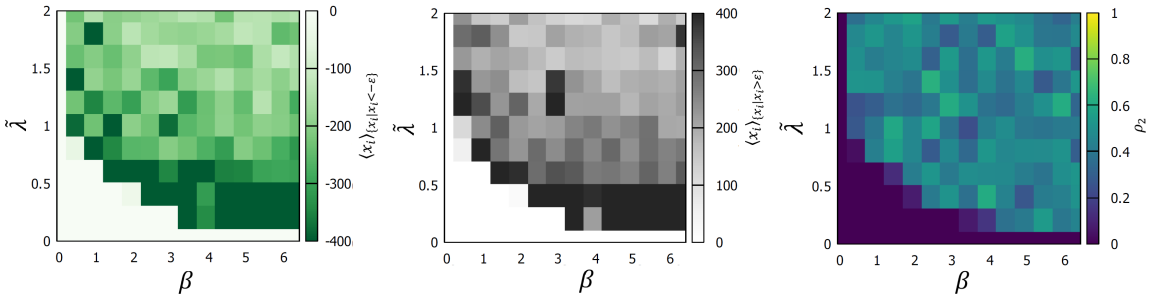
$$\rho_2 = \frac{|\langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f > \epsilon\}} - \langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f < -\epsilon\}}|}{\langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f > \epsilon\}} + \langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f < -\epsilon\}}} \tag{24}$$

If the final state was completely radicalized, then $\rho_2 \approx 1$. If the state had a completely polarized state with symmetric modes, $\rho_2 \approx 0$. Exclusively in those limit cases fulfil the condition $\rho_1 = \rho_2$. In the rest of the possible states, the value of ρ_2 is usually different from ρ_1. In Eq. [21] we defined ρ_1 as an indicator of the fraction of the number of non-consensus opinions with one specific sign $\sigma(x_i)$ (the maximum fraction between both signs). However, ρ_2 measures how large are is the value of the mean of the positive and negative opinions. The larger ρ_2, the more extreme the mean of the opinions is.

As we explained in Section 5.1.2, we are not able to undoubtedly describe the systems' state using only ρ_1. This way, playing with both ρ_1 and ρ_2 we will be able to characterize the state of the system as discussed in 5.2.

5.2 Results

We already clearly defined the difference between consensus and radicalization or polarization using ρ_0 as we see in Fig.8(a). However, analysing the parameter ρ_1 in Fig.8(b) by itself we realised that we were not able to identify it's state. We can also study as seen in Fig.9 the measure ρ_2 by itself and the main parameters that define this measure: the negative opinions' mean $\langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f < -\epsilon\}}$ and the positive ones $\langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f > \epsilon\}}$.



(a) $\langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f < -\epsilon\}}$ (b) $\langle \hat{x}_f \rangle_{\{\hat{x}_f | \hat{x}_f > \epsilon\}}$ (c) ρ_2

Figure 9: In Fig.9(a) and 9(b) we show respectively how large is the mean of the negative and positive opinions. We can see how in both figures, close to the consensus zone (white), we have the most extreme opinions' means (dark green and black) as we explained in Fig.7. In Fig.9(c) we can observe how ρ_2 never reaches, p.e. 1 (completely polarized state) despite it reaching $\rho_2 \approx 0$ for consensus or for highly polarized states.

However, for every enforcement of the algorithm, using the sames values for β and $\tilde{\lambda}$, we obtained completely different results, for these reason. For this reason, we decided to study the evolution with $\tilde{\lambda}$ of the final state of the system for a fixed value of β , and vice-versa. Now, in order to attain

a better comprehension of the opinions' final states of the system, we will analyze the model's behavior of each realization using simultaneously the information given by the measure ρ_1 and ρ_2 at each simulation. Through this process we were able to identify four qualitatively different dynamical regimes regarding the distribution of the final opinions $\hat{x}_f$:

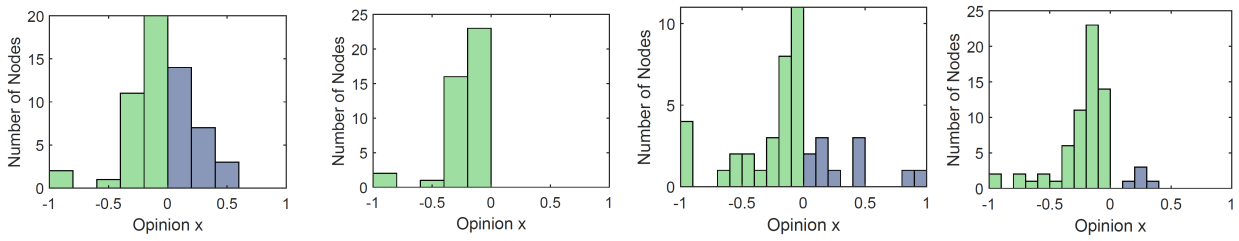
- Consensus: For small values of $\beta, \tilde{\lambda}$ there is a convergence of $\hat{x}_f$ as the only surviving opinion is the neutral one. This can be seen in the opinion distribution as one remarkable peak on the neutral state $\hat{x}_f = 0$ as we saw before. This state was clearly defined by the measure ρ_0.

- Polarization: Initial opinions acquire more extreme values far away from the neutral state, both with positive and negative sign. Unlike results from the model presented in Section 3.3 at the final state. Essentially only two opinions survive but opinions in the consensus also survive. This leads to three modes in the $\hat{x}_f$ distribution (Fig 10(a)). We obtain radicalised state for $\rho_1 \approx 0$ and $\rho_2 \approx 0$, they not usually reach 0 as in polarization states the number of negative and positive opinions is similar but not equal; and also its opinion's' means are not symmetric.

- Radicalization: the initial opinions acquire more extreme values far away from the neutral state either with positive or negative sign. We can see how the surviving final opinions are also a extreme or in the consensus. There is also a decline of surviving opinions with a opposite sign to the extreme one (Fig 10(b)). We obtain radicalised state for $\rho_1 = 1$ and $\rho_2 = 1$.

- Bias: the initial opinions acquire more extreme values far away from the neutral state either with positive or negative sign. We can see how the surviving final opinions are extreme or in the consensus. There is a decline of surviving opinions with one of the signs and the opposite sign opinion $\hat{x}_i$'s get more extreme. The opinions are biased in one concrete opinion sign, with more (See Fig.10(c)) or less biased results (See Fig.10(d)). We obtain a biased state with $\rho_2 \longrightarrow 0$ and intermediate values of ρ_1. The higher ρ_1, the more biased is the state.

In contrast with the results from Section 3.3 here we are able to distinguish a new state, the biased state, where we are able to see extreme opinions emerge but we do not necessarily obtain radicalization or polarization.



(a) Polarization (b) Radicalization (c) Biased (d) Highly biased

Figure 10: In these graphics we observe the different results of the simulations for the same parameters $\tilde{\lambda} = 1$ and $\beta = 1$. In these representation we excluded the opinions in the neutral state to facilitate the visualization of the other opinions.

6 Conclusions

The main goal of this Final Degree Project was to study the opinions' dynamics in interconnected human societies and how the interaction between individuals leads to complex social outcomes on a large scale. The primary focus was to address if the emergence of a global agreement on a given issue is originated through dynamic interactions.

Our study tries to reflect how society members "pay greater attention" to the ideas that support their own beliefs. When this partisanship is strong enough, echo chambers emerge, resulting in individuals only interacting with those whose opinions are similar to theirs. This way, they may become entangled in a concrete mindset. Furthermore, as empirical data suggests, we use both social influence and controversy as a basis for describing the rise of polarization and echo chambers.

Using dynamics that reflect the essence of echo chambers, we try to show how moderated initial opinions may evolve into a wide range of different and even extreme points of view deviating from a global consensus. We started by outlining the model proposed in the article, which characterizes social systems as complex adaptive networks where social phenomena arise from the changing structure of human relationships. Here the dynamics of each individual is affected by the pattern of connections; this pattern is encoded in the adjacency matrix that changes using an AD model [29].

The suggested model provides an overview of the emerging echo chamber phenomena and social polarization and discerns between different population states: consensus, radicalization and polarisation in social opinion. Nevertheless, despite the model's success in representing polarized states, we did not find it to be a faithful representation of reality. Firstly, one of the reasons is that we construct a new adjacency matrix at each time instant rather than time-evolving. They employ a selection probability to decide which agents—active or passive—will interact with one another at each time step as we show in Section 3.5. Secondly, in this model, the future opinions are independent of the past ones. As there is no memory of the past opinions, moderated agents can evolve into highly radical ones in a single time step. Lastly, the parameters λ and β represents the same degree of freedom.

This way we developed a new model that addresses these problems and incorporates memory into the adaptive network's evolutionary process. This adaptation is inspired by the one introduced in [37].for a network of Kuramoto oscillators. Here, the pattern of connections evolves and adds memory for the previous states. This new proposed model can replicate some empirical aspects seen in social networks as we can distinguish between consensus, polarization, radicalization and bias in social opinion. In contrast to the previous model, the final opinions' values are more realistic since some individuals are always in a neutral state. Furthermore, we are able to distinguish a new state, the biased state, for which we are able to see extreme opinions emerge but we do not necessarily obtain radicalization or polarization.

Therefore, we now have a better understanding of social states where is not only a transition between the consensus and polarized/radicalized states; we now also have a transition between consensus and extreme opinions where generally there is a bias state. Thus, polarization and radicalization are nothing other than a special case of extreme opinions.

References

[1] P. Fieguth, *An Introduction to Complex Systems* P. Springer (2016).

[2] R.N. Mantegna, et al., *An Introduction to Econophysics*, Cambridge University Press (1999).

[3] M.Keeling, and P.Rohani, Modeling Infectious Diseases, Princeton University Press, 2007.

[4] S. Galam, Y. Gefen, and Y. Shapir, Journal of Mathematical Sociology **9**, 1 (1982).

[5] S.Galam, International Journal of Modern Physics C **19**, 409 (2008).

[6] S.C.Manrubia, et al., *Emergence of Dynamical Order* , World Scientific (2004).

[7] M.E.J. Newman, *Networks: An Introduction*, Oxford University Press (2010).

[8] N.Boccara, *Modeling Complex Systems*, New York: Springer-Verlag (2010).

[9] P.W. Anderson, Science **177**, 393 (1972).

[10] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).

[11] R.K. Garrett, J. Comput. Mediat. Commun. **14**, 265 (2009).

[12] E. Dubois, G. Blank. Inf. Commun. Soc. **21**, 729–745 (2018).

[13] K. Garimella, *et al.*, Proceedings of the 2018 World Wide Web Conference, pp. 913-922 (2018).

[14] M.D. Vicario,, *et al.*, Sci. Rep. **6**, 37825 (2016).

[15] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, EPJ Data Science **8**, 35 (2019).

[16] F. Baumann, P. Lorenz-Spreen, I.M. Sokolov, and M. Starnini, Phys. Rev. Lett. **124** (2020).

[17] K. Garimella, G. Morales, A.Gionis, and M. Mathioudakis, T. Soc. Comput. Simul. **1**, 3 (2018).

[18] M. Buchanan, *The Social Atom*, Bloomsbury (2007).

[19] B.Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics, Springer (1998).

[20] R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[21] T. Gross, and B. Blasius, J. R. Soc. Interface **5**, 259 (2008).

[22] S. Boccalettia, V. Latorab, Y. Morenod, M. Chavezf , D.U. Hwanga, Phys. Rep. **424** (2006).

[23] P. Erdös, A. Rényi, Publ. Math. Debrecen **6**, 290 (1959).

[24] D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998).

[25] A.L. Barabasi, Nature (London) **435**, 207 (2005).

[26] M. McPherson, L. Smith-Lovin, and J. M. Cook, Annual Review of Sociology **27**, 415 (2001).

[27] G.G. Turrigiano and S.B. Nelson, Nat. Rev. Neurosci. **5**, 97 (2004).

[28] R.I.M. Dunbar, J. Hum. Evol. **22**, 469 (1992).

[29] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A.Vespignani, Sci. Rep. **2**, 469 (2012).

[30] A. Moinet, M. Starnini, and R. Pastor-Satorras, Phys. Rev. Lett. **114**, 108701 (2015).

[31] D. J. Isenberg, J. Pers. Soc. Psychol. **50**, 1141 (1986).

[32] Y. Kuramoto, Lecture Notes in Physics, Vol. **39**, (1975).

[33] S. H. Strogatz, Physica D: Nonlinear Phenomena **143**, 1-20. (2000).

[34] J.A. Acebrón, *et al.*, Rev. Mod. Phys. **77**, 137 (2005).

[35] J.G. Restrepo, E. Ott, and B.R. Hunt, Phys. Rev. E **71**, 036151 (2005).

[36] J. Hofbauer, *et al.*, *Evolutionary Games and Population Dynamics*, Cambridge U. P. (1998).

[37] S.Assenza, R. Gutiérrez, J. Gómez-Gardeñes, *et al.* Sci Rep **1**, 99 (2011).