GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Bachelor's Thesis

Mögliche Depolarisierungsmechanismen in mehrdimensionalen Meinungsmodellen

A Look into Possible Mechanisms for Depolarization in Multidimensional Opinion-Models

prepared by

Robin Cedric Danek

from Sondershausen

at the Max Planck Institut für Dynamik und Selbstorganisation

Thesis period:	8th August 2022 until 14th November 2022
Supervisor:	Dr. Joao Pinheiro Neto
First Rreferee:	Prof. Dr. Stephan Herminghaus
Second referee:	Prof. Dr. Stefan Klumpp

Abstract

In this bachelor thesis different ideas of depolarization mechanisms are implemented in the multidimensional opinion model created by Baumann et. al. These ideas can be categorized into manipulating the most active users of a network, implementing moderators and deleting extreme users from a network. It will be shown that in the model only the implementation of moderators and the reduction of activity lead to a recognizable effect of depolarization. At the end the application of the mechanisms of depolarization to real systems will be discussed and it will be shown that further research with different models is needed to form a conclusion of whether the different mechanisms for depolarization could work in real social systems.

Keywords: Physics, Bachelor thesis, Opinion Models, Polarization, Depolarization, Social Systems

Contents

1. Introduction

Our society is becoming more and more polarized. Whether it is the discussion about restrictions concerning Covid-19 [17], people in the USA being in the Democratic or the Republican party growing more distant from each other [16], people being for or against welcoming refugees in a country [3] or the discussion about abortion [15]. Peoples opinions on controversial topics are able to be separated into two or more clusters, meaning that people are either for or against a topic, with the people in each group not talking to people from the other groups anymore. This case is called polarization [19]. In order to bring these groups back together effective methods of depolarization have to be found.
In this thesis different ideas for depolarization mechanisms, namely using very active users, implementing moderators and deleting extreme users, are tested on a multidimensional opinion model.

2. The model

In the following the theory of the model used, the reason for using it and it's implementation are described.

2.1. Opinion Dynamics

The model by Fabian Baumann et. al. [5] is an opinion model that aims to model the formation of consensus, polarization and ideology, which is the case when polarized opinions are correlated between multiple topics, in a multi-dimensional opinion-space.
Each agent can hold a set of T opinions on T topics. These topics can be related to each other because of similar arguments that can be made when discussing them which results in the correlation of the opinions of an agent on these topics. The opinions of each agent i are represented by an opinion vector

$$\boldsymbol{x}_i = (x_i^{(1)}, x_i^{(1)}, ..., x_i^{(T-1)}, x_i^{(T)}),$$

in which the components take values from $x_i^{(\nu)} \in (-\infty, +\infty)$. The sign of each component represents the agent's stance towards the topic and $x_i^{(\nu)}$ denotes the strength of his conviction.
For the opinion of agent i to change the agent has to interact with other agents j with $j \neq i$. All connections between agents are captured by the adjacency matrix A_{ij}, where $A_{ij} = A_{ji} = 1$ means that agents i and j are connected and $A_{ij} = 0$ means that they are not [22, p.28]. The change of the opinion on topic ν of agent i is described by the ordinary differential equation

$$\dot{x}_i^{(\nu)} = -x_i^{(\nu)} + K \sum_j A_{ij} \tanh\left(\alpha [\boldsymbol{\Phi} \boldsymbol{x}_j]^{(\nu)}\right) \tag{2.1}$$

which leads to N times T ordinary differential equations describing the dynamics of a network of N nodes each having a set of T opinions. In equation (2.1) the influence of the opinions of agent j on agent i is calculated by using the $\tanh(x)$ function which is used for lessening the impact of extreme opinions. The matrix $\mathbf{\Phi}$ expresses the topic overlap of all T topics. Its entries are $\Phi_{u\nu} = \cos(\delta_{u\nu})$ where $\cos(\delta_{u\nu})$ is calculated by the scalar product of the basis vectors of topics u and ν, $\cos(\delta_{u\nu}) = \boldsymbol{e}^{(u)} \cdot \boldsymbol{e}^{(\nu)}$. This results in agent j's opinion on topic u influencing agent i's opinion on topic ν if topics u and ν are correlated which corresponds to their basis vectors not being orthogonal to each other.
The overall strength of social influence on each agent in equation (2.1) is controlled by parameter K. α is to be interpreted as the controversialness of the topics, meaning that when a topic is highly controversial $\alpha \gg 1$ in which case even moderate opinions have a big impact.

2.2. Network Dynamics

The network on which the model works is not static but instead temporal. This means that after each iteration all edges between agents are cut and new links are formed. Because of this the adjacency matrix is time-dependent, $A_{ij} = A_{ij}(t)$.
Each agent has a certain activity a_i which corresponds to his probability to become active in each iteration. The activity distribution follows a power-law distribution [23] with exponent $-\gamma$. If an agent is activated he contacts m other agents. The probability of agent i contacting agent j is dependent on their opinion distance

$$p_{ij} = \frac{d(x_i, x_j)^{-\beta}}{\Sigma_j d(x_i, x_j)^{-\beta}} \tag{2.2}$$

where β controls the homophily [21] of the network meaning that for large β agents mostly connect to agents that are similar to themselves (strong homophily) and for small β agents are more likely to connect to others that are not similar to themselves compared to the case of strong homophily. Thus small β represent the case of weak homophily. The distance d between two agents is induced by the scalar product

$$\boldsymbol{x}_i \cdot \boldsymbol{x}_j := \boldsymbol{x}_i^T \mathbf{\Phi} \boldsymbol{x}_j = \sum_{u,\nu} x_i^{(u)} x_j^{(\nu)} \cos(\delta_{u\nu}) \tag{2.3}$$

resulting in the distance of two agents being

$$d(x_i, x_j) = \sqrt{(\boldsymbol{x}_j - \boldsymbol{x}_i) \cdot (\boldsymbol{x}_j - \boldsymbol{x}_i)}. \tag{2.4}$$

which is the Euclidean Distance. When a connection between agents i and j is established at time t the adjacency matrix is updated to $A_{ij}(t) = A_{ji}(t) = 1$. A formation of a connection between agents i and j is either established by i contacting j or i being contacted by j.
After the network for an iteration was created the opinion dynamics as described in the former section take place.

2.3. Why This Model?

What makes the Model interesting for trying out different depolarization mechanisms is the number of its features. If one modifies a single feature of the model the overall functionality of the model remains the same, making it rather stable against changes.
Another important thing is the activity-driven network that´s underlying to the model. The fact that the network dynamics are determined by a probability of agents connecting makes it easier to place disturbances in the network, which could for example be an influencer. One doesn't have to heavy handedly form connections between the users and the influencer but rather the influencer connects to the network and agents with similar opinions because of the activity-driven network.

2.4. Implementation of the Model

In the following the implementation of the model in Python 3.7 is presented for the two dimensional case. At first an empty array with N rows and 3 columns is created. Each entry of the array represents an agent where the first two entries of the agent represent her opinion on two distinct topics and the third entry is her probability to become active.
Now each agent‘s entries are initialized. The opinion on each topic is drawn from a Gaussian with a mean of 0 and a standard deviation of $\sqrt{2.5}$ and the activity is drawn from a power-law distribution with exponent $-\gamma$. Since a power-law distribution with a negative exponent is undefined at zero the activities are not sampled

from $[0, 1]$ but rather from $[\epsilon_1, 1]$ where $0 < \epsilon_1 < 1$. The activities and the starting opinions of the agents are saved.
Now the network dynamics are going to be simulated. This happens in a while-loop that goes on until a predetermined number of iterations is reached.
As a first step the network for the current iteration is created. Here an adjacency list A of dimensions $N \times N$ is initialized which is filled with the placeholder value $N+1$. Also a counter-array C of dimension N is created to access the adjacency list. Now a for-loop iterates over each agent and an equally distributed random number on $[0, 1]$ is drawn for each agent. If this random number is smaller than the agent's activity the agent is activated and connects to m randomly picked agents. These agents are picked with a probability determined by equation (2.2) which is made possible by an array that includes the normalized probabilities for drawing each pair p_{ij} with $i \neq j$ that is passed on to numpy.random.choice. The connected agents are picked without replacement so that agent i can only connect to each agent once.
If agent i connected to agent j the list is updated as $A[i][C[i]] = j$ and $A[j][C[j]] = i$ and the counter array is updated as $C[i] = C[i] + 1$ and $C[j] = C[j] + 1$. This way the adjacency list doesn't only save which agents an agent actively connected to but also by which agents an agent was contacted by.
After the network is formed the opinion dynamics take place by numerically integrating equation (2.1) via the classic Runge-Kutta method [4, p.197]. Since here an adjacency list is used instead of an adjacency matrix the sum in the differential equation doesn't summarize over all agents but rather over all entries of the adjacency list A that are not $N + 1$.
For community detection one needs to retrieve a time integrated network from the temporal network. This can be done by forming an integrated adjacency matrix over a certain last number of iterations of the network. In order to do this one has to create a matrix B with dimensions $N \times N$ that is filled with zeros. When reaching a predetermined iteration of the network B can be updated from that iteration on by going over each iteration's adjacency list A and updating B with $B_{ij} = 1$ if agents i and j were connected in the network. Thus B saves all connections established within the last iterations of the activity-driven network so that now a Graph can be created from B on which community detection can be performed on.
After completing the network dynamics the last opinions of all agents are saved to a csv-file together with the activities and starting-opinions of the agents. Optionally B can be exported to a csv-file as well.

3. Reproduction of the Model

After implementing the model as described above the results of the original paper were reproduced. The reproduction is presented in this chapter.

3.1. Results of the Reproduction

At first the three different final states of the opinion dynamics the model can produce, namely consensus, polarization and ideology, were reproduced. For this the parameters were chosen as they were in the original paper [5, p.5]: $\alpha = 0.05$ $\delta = 0.0$ for consensus, $\alpha = 3.0$ $\delta = 0.0$ for polarization and $\alpha = 3.0$ $\delta = \cos(\pi/4)$ for ideology and $\gamma = 2.1$, $K = 3$, $N = 2500$, $m = 10$, $T = 2$ and $\epsilon_1 = 0.01$ as the remaining model parameters. The opinion's probability density functions (PDF)[1] and the distributions of agents in the opinion-space are shown in figure 3.1.

[1] Quelle

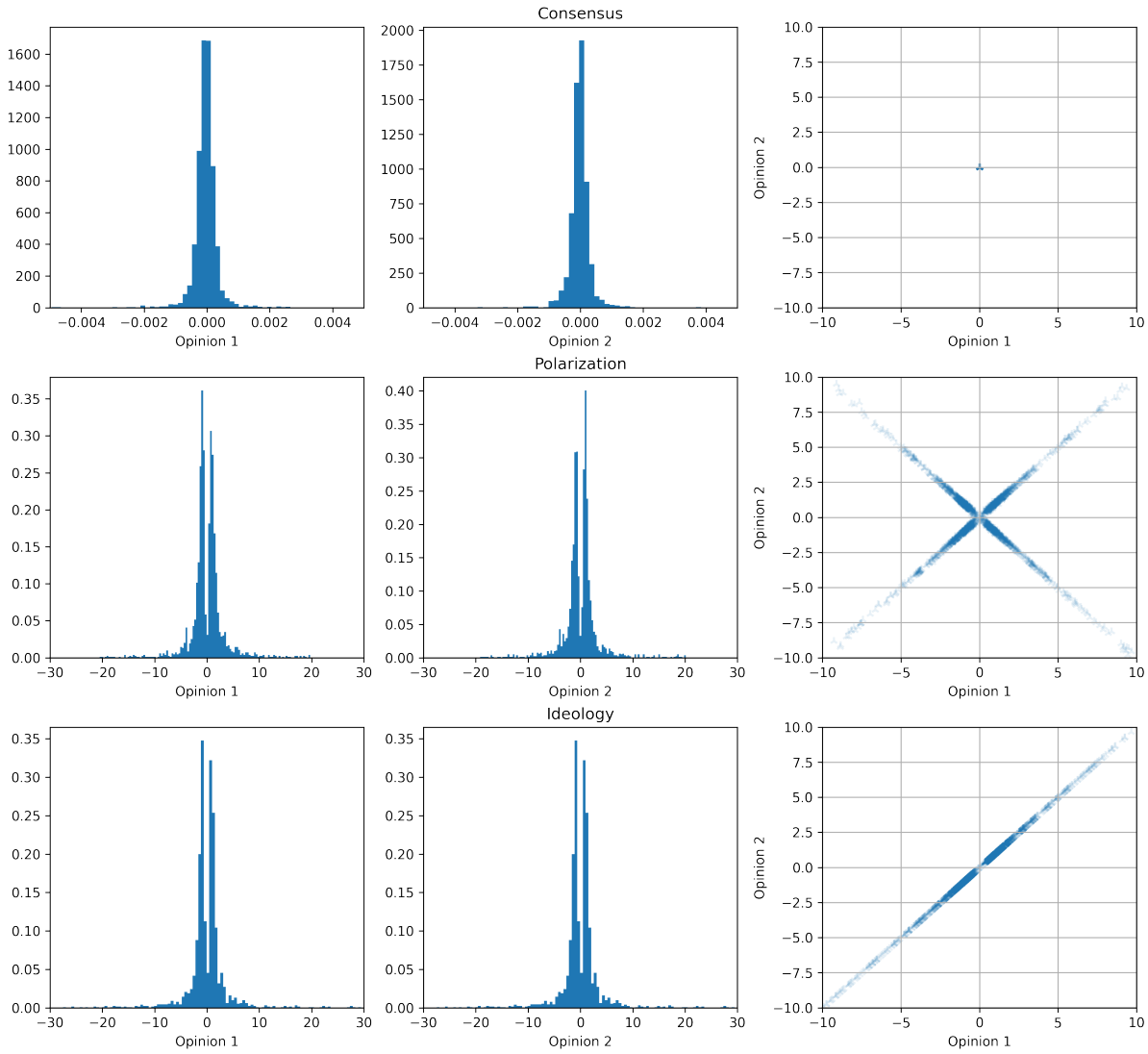

Figure 3.1.: In this figure the PDF of agent's opinions towards two topics and their placement in opinion-space are shown for the cases of consensus, polarization and ideology. The first two columns of figures show the PDFs of the agent's opinions towards topics one and two and in the third column the distributions of agents in opinion-space can be seen. In the plots of the opinion-space the nodes have a certain transparency so one can see where agents are concentrated.

For the case of consensus with no overlap of the two topics ($\delta = 0.0$) and low controversialness ($\alpha = 0.05$) one can clearly see a single peak in which all agents are concentrated. In the opinion-space on the top right of figure 3.1 one can also

see that all agents end up very close to zero meaning that they converged towards a neutral position towards both topics.
In the case of polarization agents are concentrated around two peaks for each topic. One can thus see in the opinion-space that most agents end up taking non-neutral opinions towards both topics with nodes being able to have a positive stance towards one opinion but a negative one towards the other topic at the same time which distinguishes the polarized case from the ideological case.
In the ideological case the distribution of agent's opinions towards the topics is similar to the case of polarization but looking at the opinion-space one can see that agents are either positive or negative towards both topics, which indicates that ideology is present.
Now that each final state was reproduced the parameter-space of the mean-field approximation of the model [5, p.7] as shown in figure 3.2 had to be reproduced numerically.

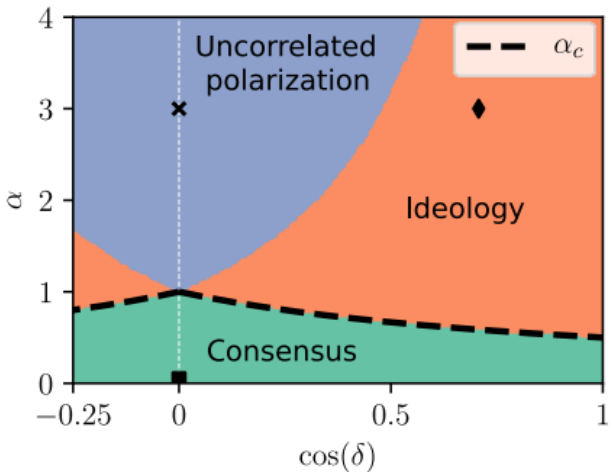

Figure 3.2.: The phase-space of the model's mean-field approximation is shown. Source: [5].

In order to do simulations that closely resemble the mean-field approximation that uses the assumptions $N \to \infty$ and $\beta \gg 1$ one has to understand how the model behaves when changing N, m and β. An analysis of the behavior of the model when changing these parameters, a stability analysis of the model and the parameter choice for all simulations can be found in Appendix A.
With the parameters being set the opinion distributions of agents had to now be classified into resembling consensus, polarization or ideology in order to recreate the

phase-space of the mean-field approximation.

3.2. Classifying Distributions

The detection of whether consensus, polarization or ideology is present in a final opinion distribution is of importance since looking at what different manipulations done to the model do to the parameter-space of its mean-field approximation [5, p.6-7] helps to understand the effect of the manipulation. In the following two classification-methods will be discussed.

3.2.1. Community Classification

The first way of classifying distributions was by looking at the communities formed in the network and their position in the opinion-space. Since the network of the model is a temporal one and since only active nodes connect to a small number of other agents community detection on the network formed in the last iteration would only lead to some small communities and many nodes forming a community on their own.
In order to find meaningful communities the approach of Baumann et. al. [5, p.8-9] was followed. Here an integrated network was formed by saving all connections made over the last 70 time-steps of a simulation, as was already described in the former section. On this integrated network communities can be detected by using the Louvain algorithm [20, p.3-4].
Once communities were detected the ones that were too small, namely communities with a number of nodes smaller than $N/1000$, were filtered out. The mean angle of nodes being in a community was then calculated for each community. It was then checked whether these mean angles took values of 45°, 135°, 225° or 315°, each with a tolerance of ±30°. If the mean angles were distributed over all 4 angles the distribution was classified as polarization. If the mean angles were found to be distributed only around 45° and 225° or 135° and 315° the opinion distribution was classified as ideology. In the case of not all mean angles being able to be categorized as being close to one of the four angles the distribution was classified as consensus. For reproducing the phase-space of the mean-field approximation with this classification method three simulations were classified per parameter pair for a parameter range of $\alpha \in \{0, 0.1, ..., 3.9, 4.0\}$ and $\cos(\delta) \in \{0, 0.1, ..., 0.9, 1.0\}$. The results can

be seen in figure 3.3. While there is some resemblance between the two phase-space plots the numerically calculated one is prone to a lot of fluctuations and wrongly classifies many polarized states as consensus and many consensus states as polarized or ideological.

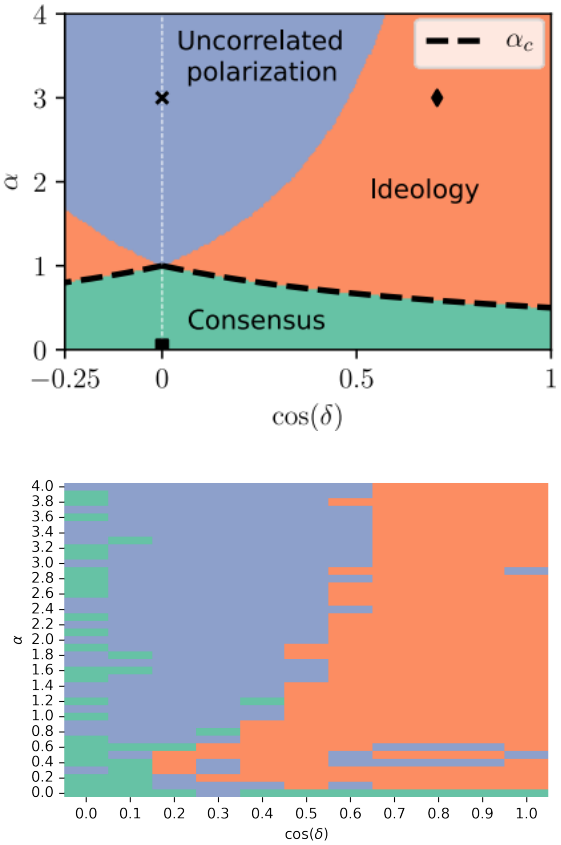

Figure 3.3.: In the top figure the analytically calculated phase-space of the model can be seen. Below is the numerically calculated phase-space which was created using community classification.

3.2.2. Agglomerative Classifier

An improvement to the Community Classification is a classifier based on an agglomerative algorithm designed to find clusters of nodes. This algorithm is based on the idea of merging agents to one agent if their opinion difference is below a certain threshold and then creating a new agent with his opinion vector being a weighted mean of the opinion vectors of the former agents. This newly created node has an increased weight compared to his parent-nodes causing opinion-vectors of nodes that

emerge from him and another node to be close to him.
Again for a two-dimensional topic-space an array of the size $N \times 3$ is created. The first two entries are the opinions of an agent which are drawn from the final opinion distribution that is to be analyzed. The third entry is initialized with 1. This entry resembles the number of nodes that are already merged within a node.
The algorithm starts with a while-loop that goes on until a certain runtime is reached. At the beginning of each iteration, an array I containing all normal numbers from 0 to $N_t - 1$ is created, with N_t being the number of nodes that remain at iteration t. Now a for-loop loops over all agents and it is checked whether the current agent is included in I. If it isn't the agent is skipped.
For each agent another agent is picked randomly where now the probability of picking an agent is uniform for all agents and it is checked whether their opinion distance, which is the Euclidean distance, is below a treshold η. This threshold should be roughly equal to the absolute position of a peak of opinions in a polarized distribution. If the distance is smaller than η the nodes are merged and a new node is created. Its position in the opinion-space is calculated by

$$\vec{o} = \frac{n_i \cdot \vec{o_i} + n_j \cdot \vec{o_j}}{n_i + n_j} \tag{3.1}$$

where n_i is the weight of node i being the number of nodes that were merged within i. Because of this equation nodes that consist of many nodes don´t move much anymore when merging with single nodes but rather draw them towards their position resulting in the "heavy" nodes sitting roughly where the peaks of the opinion distribution are at if η is chosen correctly.
After creating the new node the old two nodes are deleted and their numbers are cleared from I. When the for-loop ends the remaining nodes that were not merged and the newly created ones are put together in a new array. For the new array $N_t \leq N_{t-1}$ is always the case. Thus N_t converges until only nodes are left that can't be merged since their distance is bigger than η.
With the final nodes produced by the agglomerative algorithm one can classify the distributions into consensus, polarization and ideology. For this two things are of importance: The angle at which the final nodes sit at in the opinion space and how far they are from the center. The angle here is the angle between an opinion vector and the x-axis.
At first each final node is categorized into either having an angle of 45°, 135°, 225° or

315° again with a tolerance of ±30°. After that the final nodes are categorized into indicating polarization or consensus by putting them into the consensus-category if all their opinion's absolute is below a threshold θ and into the polarization-category else.
Now it is checked whether more nodes are in the consensus-category or in the polarization-category. This is done by summing up all nodes merged within the final nodes in each of the two categories and comparing the amounts. If more nodes are within the consensus category the distribution is classified as consensus. If that isn´t the case, the angles are checked. If there are final nodes only in either 45°, 135°, 225° or 315° the distribution is classified as ideology. If there are nodes in each angle-area the distribution is classified as polarization.
Since the final nodes are randomly formed and since their positions and node distributions of the final nodes are prone to fluctuations multiple runs of the agglomerative algorithm are done for each data set and the state that was classified most often is chosen as the state of the data set. If two states were classified equally often consensus is chosen over polarization and ideology and polarization is chosen over ideology.
For simulations with $m = 10$ the peaks of a polarized state end up around an absolute value of one for each opinion, thus $\eta = 1.0$. Through numerical analysis $\theta = 0.5$ was found to lead to the best results. Figures depicting example runs and their classification can be found in the Appendix (B)
In figure 3.4 a comparison of the analytical phase-space and the one found by using the agglomerative classifier can be seen. The agglomerative classifier is much less prone to fluctuations than the community classifier. What is important is that the agglomerative classifier finds much more consensus than what is expected from the analytical result. This discrepancy can be explained by polarization and consensus not being well defined. Especially at the phase-transition from consensus to polarization or ideology it is up to discussion whether a distribution having two peaks but many nodes between them is called consensus or polarization [19].
To further elaborate why the agglomerative classifier finds much more consensus than the analytical result different PDFs of the agent's opinions around the phase-transitions and their classification were plotted in figure 3.5. The according opinion-spaces can be found in figure 3.6. As can be seen in the figures consensus is reached for at least $\alpha \leq 1.5$ and $\cos(\delta) = 0.0$. For $\alpha = 1.8$ two peaks are visible in the PDF but there are still many nodes between them. Thus the distribution can be seen

as both polarized and consensus-like, depending on what criterion one uses. When increasing α or $\cos(\delta)$ the peaks grow more elaborate which makes the classification into polarization, consensus or ideology more clear.
The difference between the analytical and numerical results can also be traced back to $N \to \infty$ and $\beta \gg 1$ and thus the conditions of the mean-field approximation not being fulfilled perfectly since simulations were performed for $N = 2500$ and $\beta = 5$. It is possible that for bigger N the peaks near the phase-transition would be formed more clearly which would result in the classifier classifying polarization for lower values of α than it currently does.

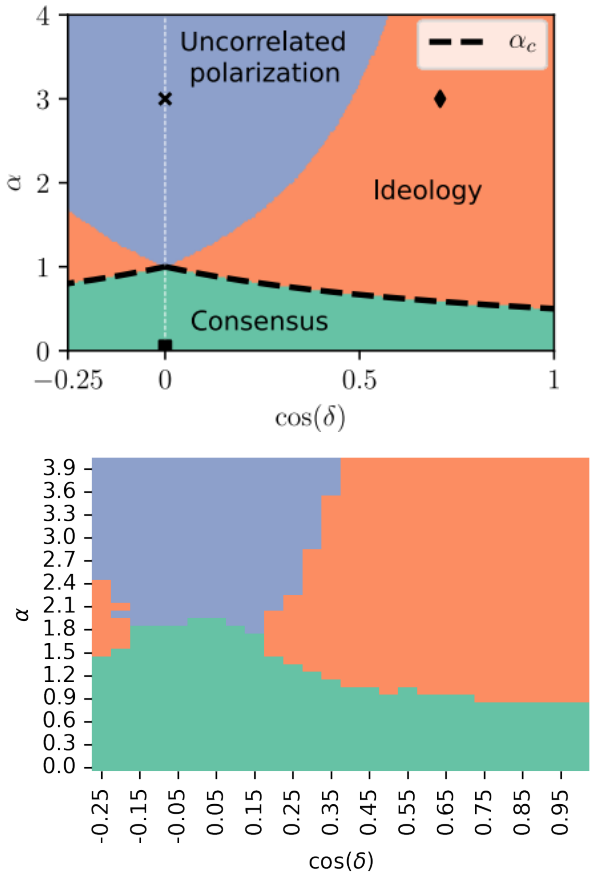

Figure 3.4.: In the top figure the analytically calculated phase-space of the model can be seen. Below is the numerically calculated phase-space which was created using the agglomerative classifier.

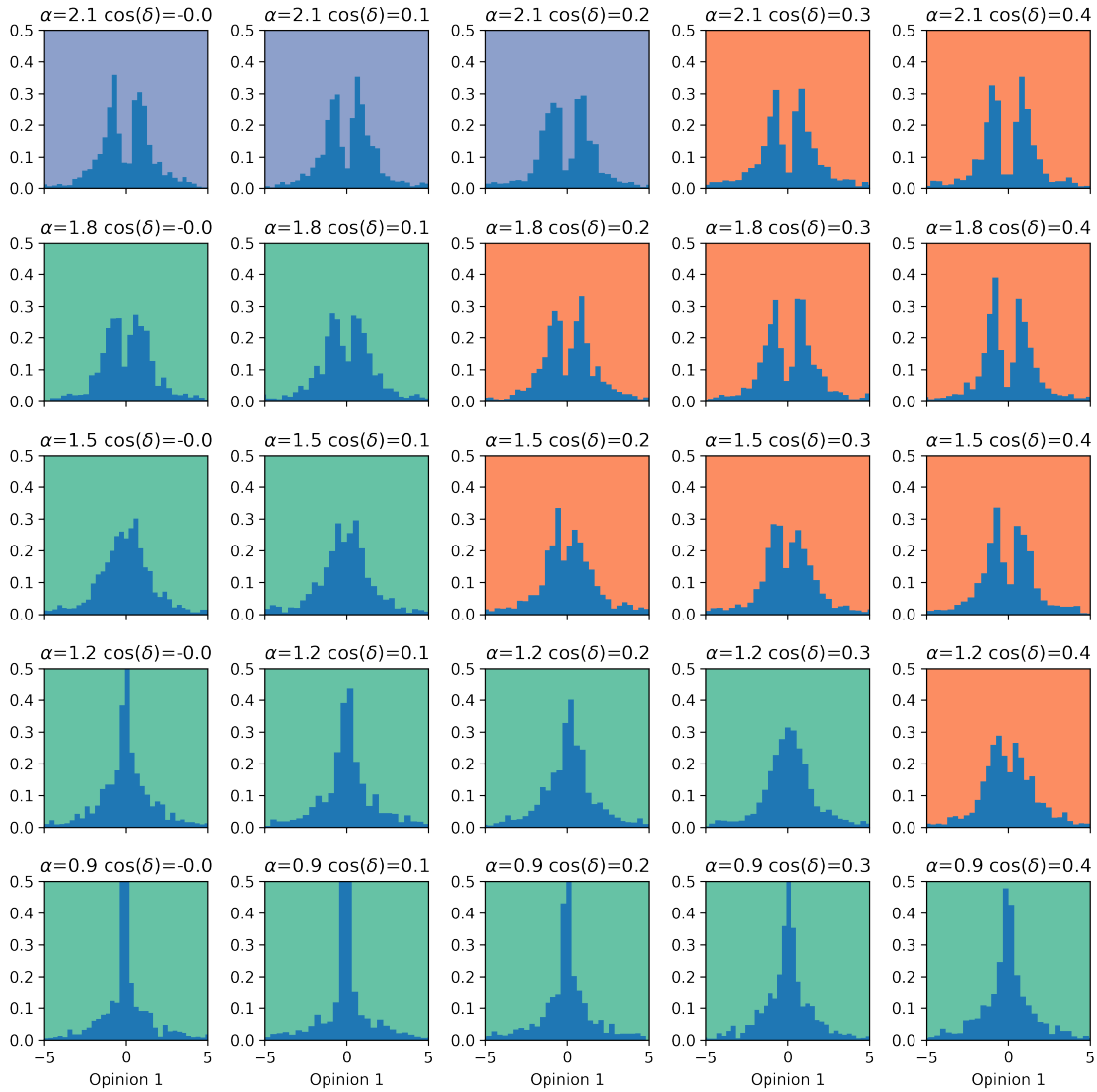

Figure 3.5.: Different resulting PDFs of opinions for diverse values of α and $\cos(\delta)$ can be seen. Their background is colored according to how they were classified by the agglomerative classifier, namely green meaning consensus, blue representing polarization and orange standing for ideology.

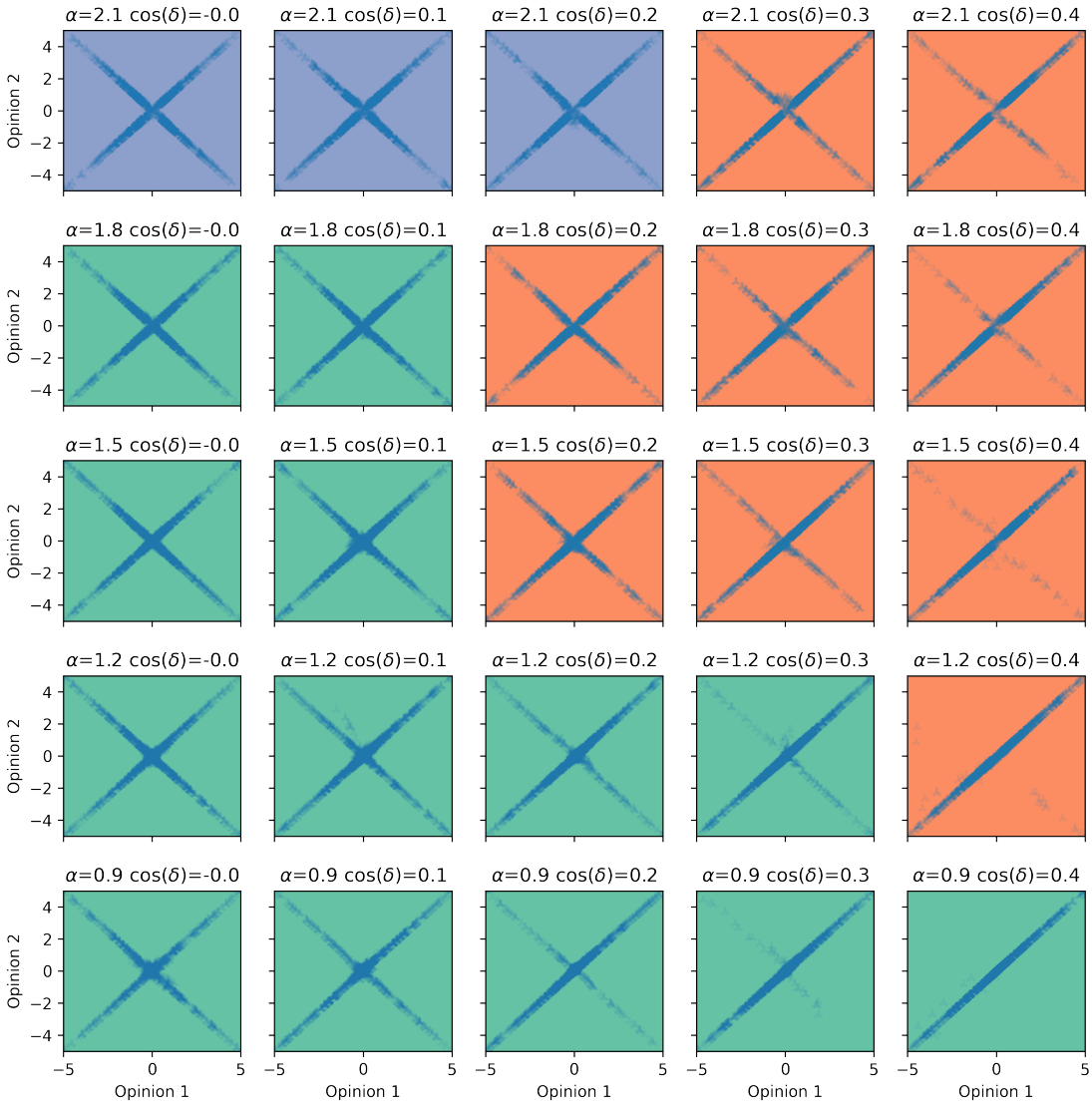

Figure 3.6.: Different final opinion-spaces for diverse values of α and $\cos(\delta)$ can be seen. Their background is colored according to how they were classified by the agglomerative classifier, namely green meaning consensus, blue representing polarization and orange standing for ideology.

4. Approaches to Depolarization by Manipulating the Most Active Users

In this section approaches to depolarization that are based on manipulating the most active users of a network will be introduced and the resulting phase-spaces and opinion distributions are going to be presented.

4.1. The Role of Super-Active Users

What motivates the manipulation of the most active users of a network? To answer this question the positions of agents in a polarized state are plotted with color-coded activity. This can be seen in figure 4.1.
What is visible is that the more active an agent is the more extreme his opinions are. This can also be seen in figure 4.2 where the absolute opinion of the agents from a network with $N = 2 \cdot 10^4$ are plotted against their activity. A linear dependency can be seen where the parameters of the fit $f(x) = a \cdot x + b$ are $a = 58.609$ and $b = 0.0411$.
Since agents influence each other in an attracting manner a possible dynamic of the network is that the most active agents of the network that sit at the extremes of the opinion spectrum pull other agents towards them which would make them very important to polarization.

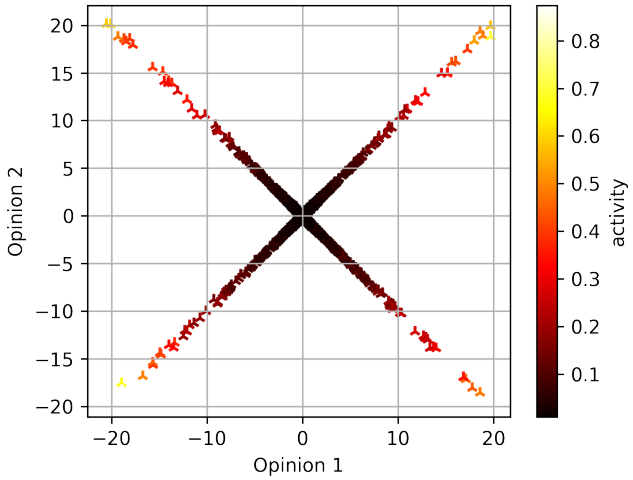

Figure 4.1.: The positions of agents in the opinion-space of a polarized state ($\alpha = 3.0$, $\cos(\delta) = 0.0$) is shown. The agent's activity is color-coded by the colorbar to the right. The state is the same as the one used in Figure 3.1.

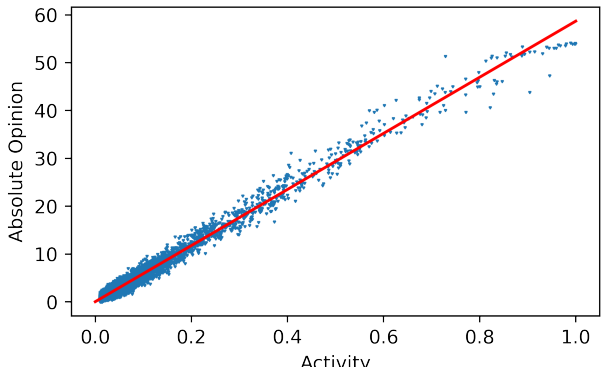

Figure 4.2.: The absolute opinions towards a topic of agents from a polarized network ($\alpha = 3.0$, $\cos(\delta) = 0.0$) of size $N = 2 \cdot 10^4$ are plotted against their activity. The data was fitted with a function $f(x) = a \cdot x + b$ with $a = 58.609$ and $b = 0.0411$ which can be seen in red.

When looking at the flow of the most active agents in a polarized network, which can be seen in figure 4.3, one can see that the most active agents move towards the extremes very quickly and that they would, according the hypothesis proposed above, exert the pull on other agents soon after starting a simulation.

If they really do play that role, manipulating their interaction strength with other agents by reducing the influence the highly active agents have on moderately active agents should lead to less polarization. This is what the following two methods are going to do.

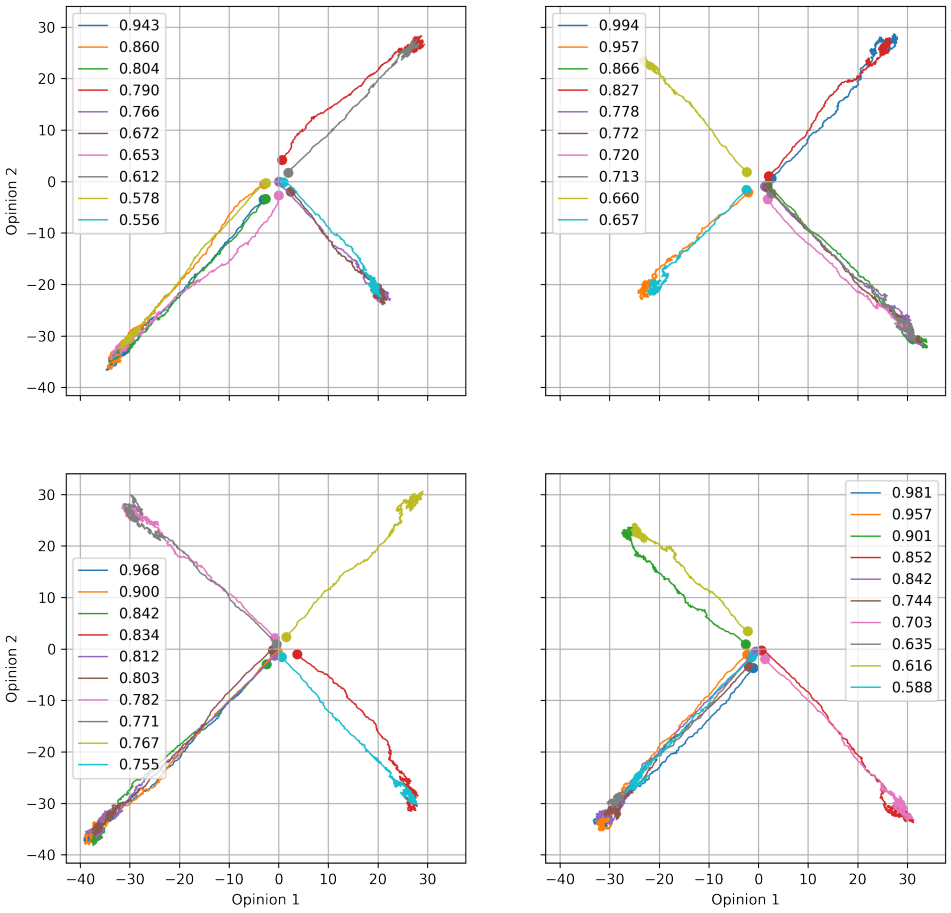

Figure 4.3.: The flow of the 10 most active agents of four polarized networks in the opinion-space is shown. Their starting positions are indicated by the points. The legend depicts the activity of the agents.

4.2. Penalizing by Activity

A first possibility of reducing the influence of the most active agents is by directly reducing the influence they have on others. This can be done by modifying equation (2.1) to

$$\dot{x}_i^{(\nu)} = -x_i^{(\nu)} + K \sum_j A_{ij} \tanh\left(\alpha [\mathbf{\Phi} \boldsymbol{x}_j]^{(\nu)}\right) \cdot f(a_j). \tag{4.1}$$

Note that for agent i only the activity of agent j is relevant for the penalty function. This is important for a very active agent j only having less influence on agent i with low activity but not vice versa. $f(a)$ can be any function for which $f(a) \in [0, 1]$ is true. Functions that have large values for low activities and small values for big activities are of interest here since they especially target highly active nodes.
In figure 4.4 both the original phase-space and the one altered with the penalty function can be seen. Here $f(a)$ is a Heaviside function [10, p.160] that is zero for $a > 0.2$ and one else. Thus $f(a)$ cuts off all of the influence the most active agents have on the network. When comparing both phase-spaces it can be seen that there is no difference between them aside from fluctuations.

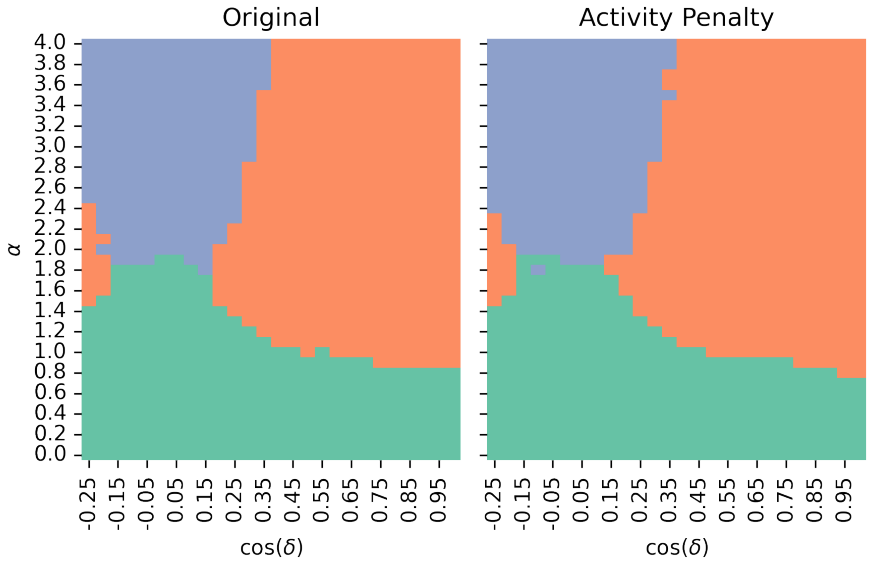

Figure 4.4.: Two phase-spaces are depicted. The one on the left is for the unchanged model and the one on the right is for the model including an activity penalty.

4.3. Penalizing Recent Connections

A different approach to penalizing highly active agents is by penalizing recent connections with other agents. Because of homophily the highly active agents should mostly be connected to a subgroup of the network. If one now penalizes repeating connections between highly active agents and agents of this subgroup the influence of the former should be lessened.
There are two ways to go about this: The first one is directly penalizing the influence active agents have on others which is similar to what has been presented in section 4.2. The second one is diminishing the connection probability between recently connected agents.
Both mechanisms can be realized by introducing a time dependent penalty matrix $P(t)$. The entries

$$P_{ij}(t) = e^{-\lambda \cdot (t - t_{ij})} \tag{4.2}$$

contain a decaying exponential. t_{ij} is the iteration at which agents i and j have connected and t is the current iteration. The penalty matrix is symmetrical since $t_{ij} = t_{ji}$. For penalizing the influence agents have on each other the differential equation of the opinion dynamics becomes

$$\dot{x}_i^{(\nu)} = -x_i^{(\nu)} + K \sum_j A_{ij} \tanh\left(\alpha [\mathbf{\Phi} \boldsymbol{x}_j]^{(\nu)}\right) \cdot (1 - P_{ij}(t)). \tag{4.3}$$

In a similar manner the penalty matrix can be used for reducing the connection probability between recently connected agents. The new connection probability becomes

$$p_{ij} = \frac{d(x_i, x_j)^{-\beta} \cdot (1 - P_{ij}(t))}{\Sigma_{k \neq i} d(x_i, x_k)^{-\beta} (1 - P_{ik}(t))}. \tag{4.4}$$

In figure 4.5 three different phase-spaces for the original model, the one with the penalty matrix reducing the influence of recent connections and the one with the penalty matrix impacting the connection probability can be seen for $\lambda = 0.1$.
When comparing both of the manipulated phase-spaces to the original one again no notable difference can be found aside from minor fluctuations. As neither penalizing recent connections or reducing the influence of highly active users notably changes the phase-space the last possible mechanism of depolarizing by manipulating the activity of users is by changing the distribution of activity in the network.

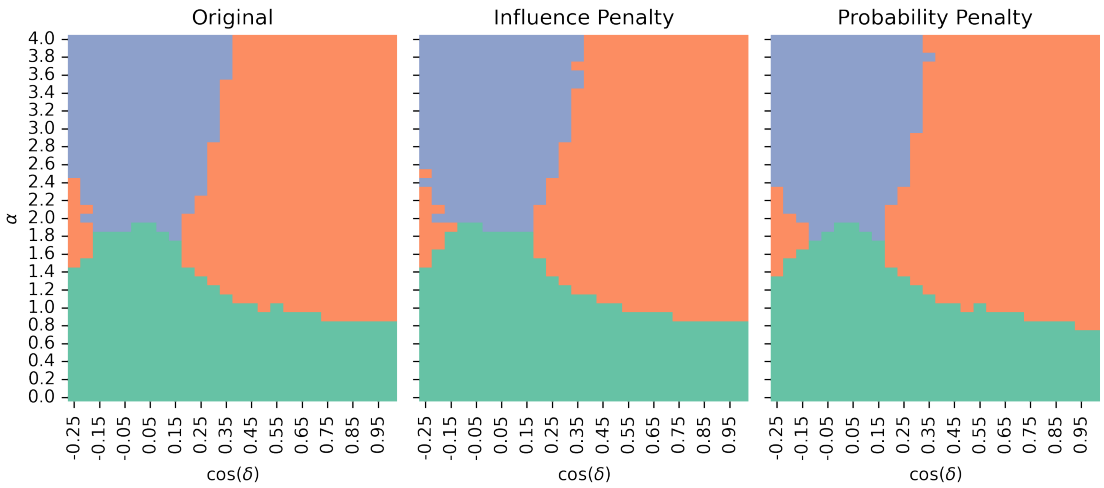

Figure 4.5.: Three different phase-spaces are shown. The one on the left was produced by the original model, the middle one was created using the model with penalized influence of recent connections and the rightmost one was produced by the model with lessened probability of recent connections happening again.

4.4. Changing the activity distribution

The last approach of manipulating the most active agents of a network is manipulating the activity distribution from a power-law with negative exponent to a Gaussian distribution [26]. This results in more agents being more active and to less agents causing most of the network's activity.

What's important here is to cap the total number of agents that can be active in a given simulation. If one simply changes the activity distribution to a Gaussian more agents will be active compared to the power-law case. This leads to stronger polarization since the term for social influence in equation (2.1) grows more dominant. Thus agents have to have a stronger absolute opinion in order to compensate the bigger social influence and to reach a fixed point [9, p.42]. Because of that the mean number of people being active in each iteration of a network of size N that has a power-law distributed activity has to be numerically determined in order to only evaluate the effect of changing the distribution of activity. This mean will be the number of agents that get activated in each iteration in a Gaussian network where

the probability to be activated is distributed in a Gaussian manner.
In figure 4.6 the phase-spaces of both the original model and the one with Gaussian distributed activity can be seen. The Gaussian activity distribution was set to have a mean of 0.5 and a standard deviation of 0.1. The mean-number of users that were active per iteration for the power-law distributed case with $\gamma = 2.1$ and $N = 2500$ was estimated over 10 simulations each with a runtime of 1000 iteration to be roughly $\bar{a} = 102$. Thus 102 agents were activated in each iteration.
When comparing the two one can see that the amount of consensus reached in the phase-space is vastly reduced for the Gaussian activity distribution. There is more ideology and even more polarization present even for small values of controversialness with $\alpha < 1.0$. Thus changing the activity distribution to a Gaussian does not lead to depolarization but rather increases polarization.

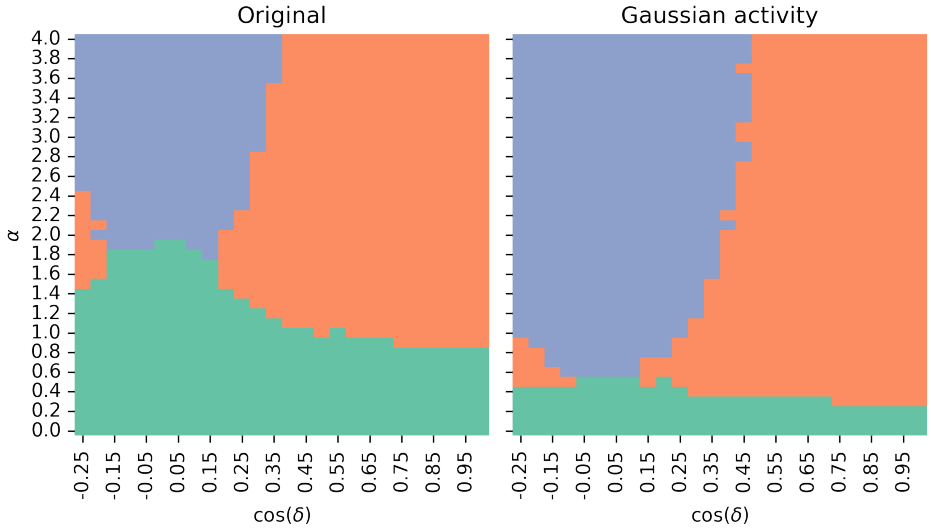

Figure 4.6.: Both the phase-spaces produced by the original model and the one with Gaussian distributed activity can be seen. The Gaussian activity distribution had a mean of 0.5 and a standard deviation of 0.1

5. Depolarizing by Implementing Moderators

The idea of the next two mechanisms was trying to depolarize by adding moderators to a network. This was inspired by the idea that moderators could possibly be implemented in social media thus helping to fight polarization without being too heavy-handed of an approach.

5.1. Moderators Causing Decay in Opinions

The first idea was creating moderators that are active in every turn ($a_i = 1$). These moderators reach more agents than normal agents ($m_{\text{mod}} \gg m$) and don't have an opinion and thus aren't bound to homophily. This results in them picking other agents either with equal probabilities or with the activity of agents being used as probability $p^{(m)}$ to be picked by each moderator with

$$p_i^{(m)} = \frac{a_i}{\sum_j a_j}$$

where $p_i^{(m)}$ is the probability of agent i to be picked by each moderator.
Once a moderator makes a connection with another agent he actively moves the agent towards a neutral opinion by adding a decay term to equation (2.1), leading to

$$\dot{x}_i^{(\nu)} = -x_i^{(\nu)} - s_m \cdot \tanh(x_i^{(\nu)}) + K \sum_j A_{ij} \tanh\left(\alpha [\mathbf{\Phi} \boldsymbol{x}_j]^{(\nu)}\right). \tag{5.1}$$

Here s_m is the strength a moderator has on the agents he connects to.
In figure 5.2 the phase-space of the original model and variations of the model including different numbers of moderators with different amounts of reaches all with a moderating strength of $s_m = 0.3$ are depicted. s_m is small compared to the average social influence an agent is exposed to for a system with high controversialness ($\alpha =$

3.0). The mean social influence was averaged over 100 simulations with $N = 2500$, $\alpha = 3.0$, $\cos(\delta) = 0.0$ and $\beta = 5.0$ and resulted in $\bar{K} \approx 0.9$. The PDF of mean social influence per agent can be seen in figure 5.1. The exponent of the power-law that is present for mean social influences bigger than 0.1 can be determined by fitting the part of the PDF of the mean social influence to $f(x) = Ax^{-\gamma_f}$. The fit results in the parameters $\gamma_f \approx 2.10983$ and $A \approx 0.06977$ with $\gamma \approx \gamma_f$ meaning that the PDF of activity and the one for mean social influence share the same exponent.

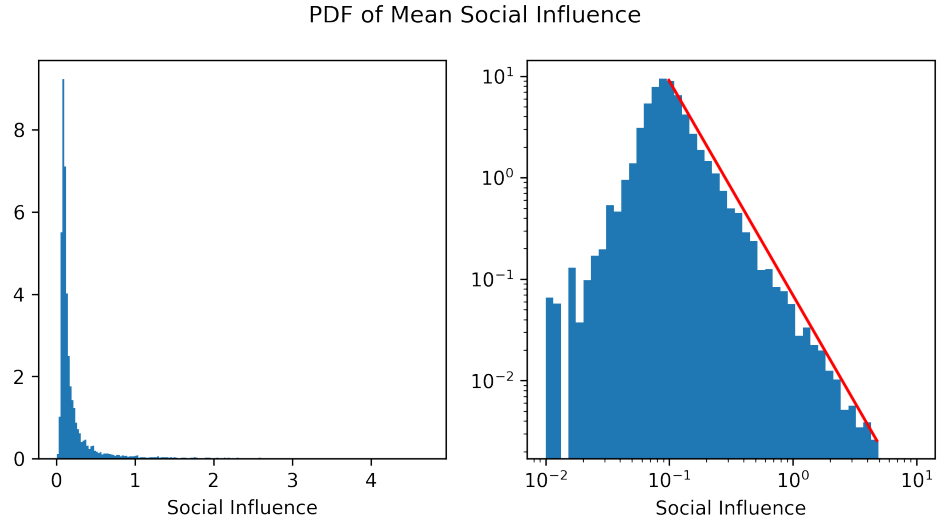

Figure 5.1.: The PDF of mean the social influence an agent is exposed to per iteration is shown both with normal scaling and with double-logarithmic scaling. The distributions were created using simulations with $N = 10^3$, $\alpha = 3.0$, $\cos(\delta) = 0.0$ and $\beta = 5.0$. The power-law $f(x) = Ax^{-\gamma_f}$ with parameters $\gamma_f \approx 2.10983$ and $A \approx 0.06977$ that was fitted to the distribution for $K \in [0.1, 4.72]$ is shown in the double logarithmic plot in red.

It is visible in the bottom phase-spaces of figure 5.2 that the added moderators with equal probabilities of connecting to agents had a depolarizing effect on the opinion dynamics, both for few moderators with high reach and for more moderators with less reach. Consensus is reached even for high controversialness with $\alpha = 4.0$ and small overlap between the topics. When increasing δ ideology emerges again but is overall less present compared to the original case. Aside from few fluctuations there is no difference between the cases of $N_{mod} = 25$, $m_{mod} = 50$ and $N_{mod} = 5$, $m_{mod} = 250$.

Evaluating the effect of moderators picking agents of the network with the probability $p^{(m)}$ by comparing the according phase-space in the top-right of figure 5.2 to the original phase-space it can observed that the moderators did have a depolarizing effect. This depolarization is less strong than the one that can be achieved by having moderators target all agents with equal probabilities.

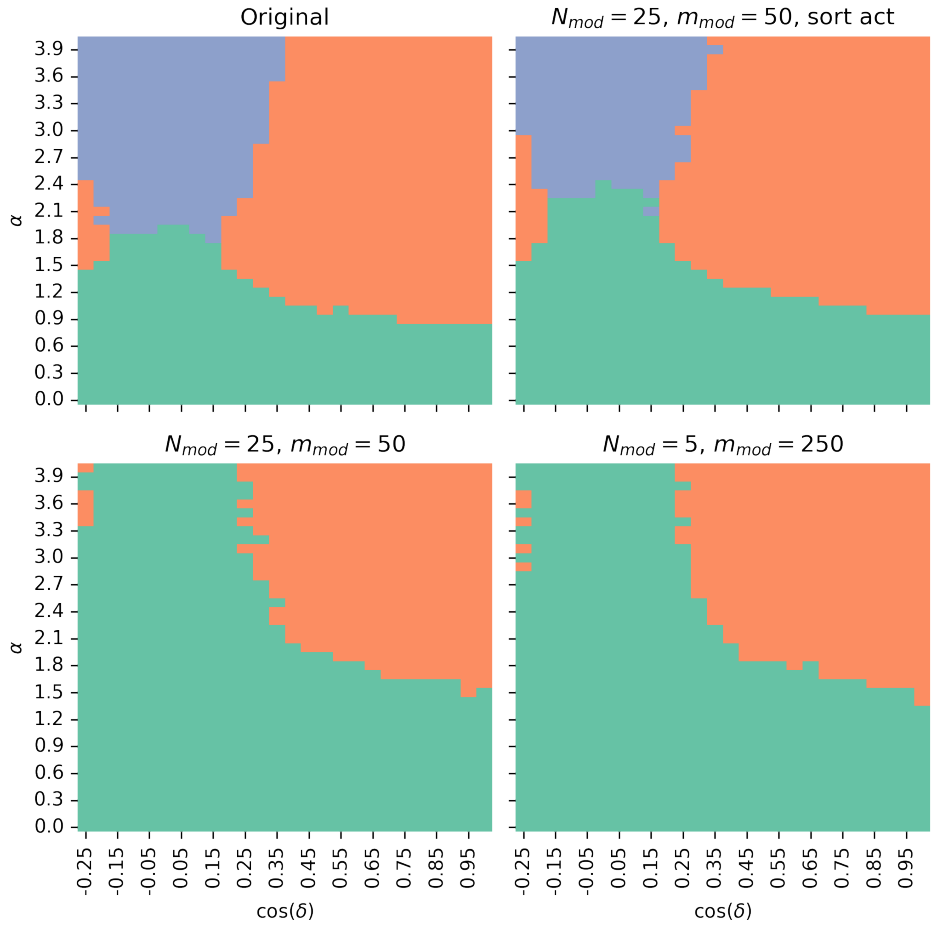

Figure 5.2.: Four phase-spaces can be seen. The top left one is the one produced by the original model and the one on the top right was created by adding 25 moderators with $m_{mod} = 50$ and $s_m = 0.3$ that picked agents with the probability $p^{(m)}$. The bottom phase-spaces resulted from implementing moderators that picked agents all with the same probability.

5.2. Moving Influencer

The second approach to depolarizing a network with moderators was by adding moving influencers that are always active to the system after it has been polarized. These influencers follow a set path through opinion-space and thus have an opinion as opposed to the moderators described in the former section. They are also bound to homophily as a normal agent is but again reach far more agents than a normal agent would. They also don't make changes to equation (2.1) but instead influence agents in the normal way. Thus through the original dynamics these influencers are supposed to pull agents towards them and move through opinion-space in order to effectively pull agents with them towards the center.
The effect on distributions with $\alpha = 1.5$ and $\cos(\delta) = 0$ for different numbers of influencers and their reaches are shown in figure 5.3. All influencers were introduced to the system at iteration 1000 and started with absolute opinions of $|x_i^{(\nu)}| = 5$. They then moved towards the center and ended up with absolute opinions of $|x_i^{(\nu)}| = 0$ at iteration 2000. The influencer's paths could be either parameterized linearly by

$$\boldsymbol{l}(x) = \begin{pmatrix} \pm 5x \\ \pm 5x \end{pmatrix} \text{ or } \boldsymbol{l}(x) = \begin{pmatrix} \pm 5x \\ \mp 5x \end{pmatrix}$$

or in a parabolic manner

$$\boldsymbol{l}(x) = \begin{pmatrix} \pm 5x^2 \\ \pm 5x^2 \end{pmatrix} \text{ or } \boldsymbol{l}(x) = \begin{pmatrix} \pm 5x^2 \\ \mp 5x^2 \end{pmatrix}$$

with $x \in [0, 1]$ which is indicated in figure 5.3 with "Lin" or "Par". The difference between the parametrizations is the velocity with which an influencers moves through opinion space. For the linear case the velocity is constant but for the parabolic case the influencer moves quickly from a strong opinion towards the center with his velocity decreasing as he moves closer towards a neutral opinion. Because of that the moving influencers spent more time in the system having a moderate opinion for the parabolic case as compared to the linear case. The influencers were always equally distributed over all four quadrants of opinion-space.
When comparing the three resulting PDFs of the agent's opinions for $N_m = 4$ and parabolic parametrization it can be seen that the bigger the impact the influencers have with increasing reach the more polarized the system becomes. The same result can be found when comparing the three final opinion distributions for $N_m = 40$.

Comparing the PDFs of the cases with parabolic parametrization for which the product $N_m \cdot m_{mod}$ is constant it can be observed that having more influencers with less reach instead of less influencers with more reach leads to stronger polarization even if the maximum number of agents that can be reached by the influencers per iteration is held constant.
The case of linear parametrization doesn't depolarize but polarizes more as well. This can be seen by the peaks of the PDF of the agent's opinions having smaller maximum values compared to the original distribution. The flattening of the peaks results from the PDFs having fatter tails meaning that some agents that would have had a moderate opinion ended up having stronger opinions compared to the original case.

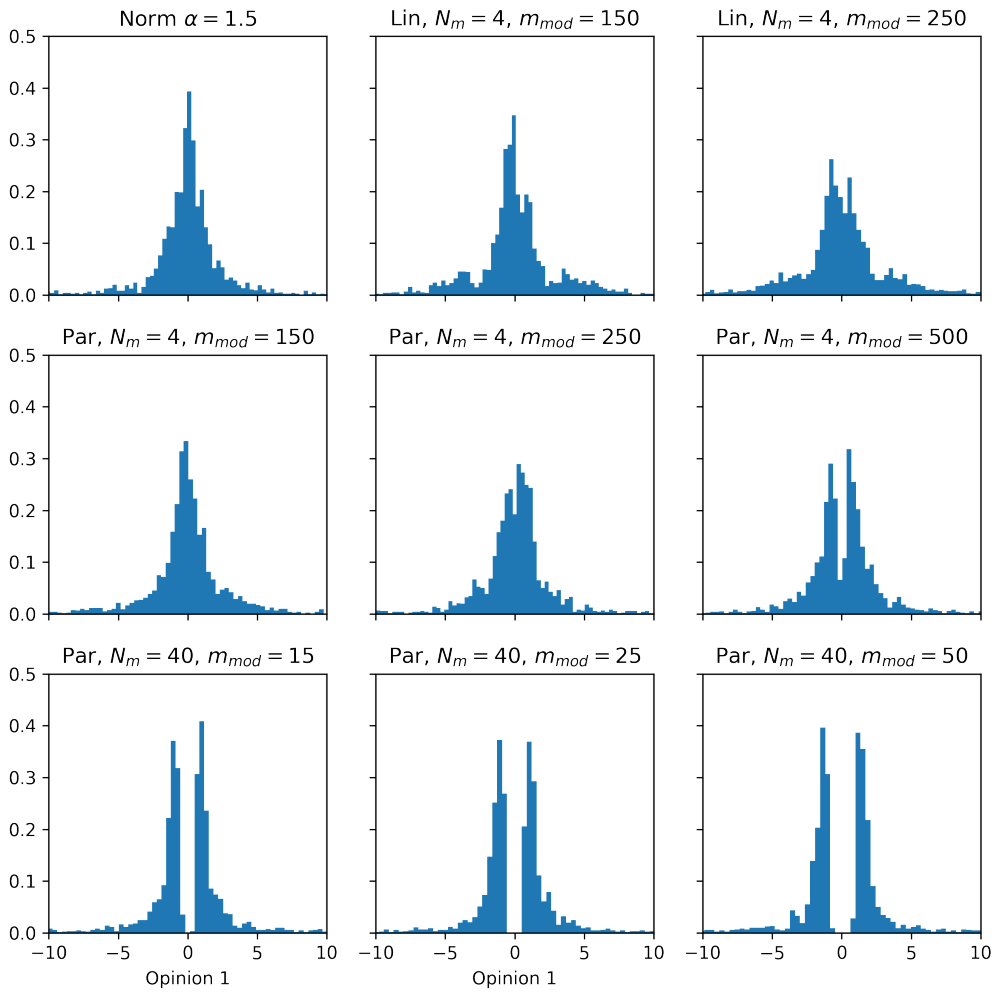

Figure 5.3.: Multiple PDFs of agent's opinions for a network near the phase-transition ($\alpha = 1.5$ and $\cos(\delta) = 0$) are depicted. The top left figure was created with the original model. The others were created with the implementation of moving influencers. N_m is the number of moving influencers and m_{mod} is their reach. "Par" or "Lin" indicates the parametrization of the paths of the moving influencers.

6. The Effect of Extreme Users Leaving the Network

A last possible mechanism that is covered is deleting users that grow too extreme from the network. This emulates users leaving a social network because of most people not agreeing with them or them being removed by moderators for having too extreme of an opinion.

This method can be implemented by checking how many users can be found within a disc that is drawn around each agent in the opinion space. Then a percentage of all users that have to be included in the disc has to be set. If the percentage of users in a disc around agent i is below this percentage limit the agent is removed and newly initialized with random opinions drawn from a Gaussian with its mean at zero and a standard deviation of $\sigma = \sqrt{0.5}$. Only the agent's opinions are newly drawn. His activity is kept the same in order to preserve the power-law distributed activity.

In figures 6.1 and 6.2 The effect of deletion on simulations with $\alpha = 1.5$ and $\alpha = 2.5$ can be seen. For both figures at least 50% of agents had to be inside a disc for a user to stay within the network. When first looking at the effect deleting agents has on a polarized distribution (figure 6.1) one can see that two peaks remain, even if more agents are affected by deletion through decreasing the radius of the disk around an agent. In the case of being close to the phase-shift (figure 6.2) two peaks form when increasing the amount of deletion in a network. For a radius of 64 the original distribution remains, but for smaller radii the PDFs are more and more polarized until they are most polarized for a radius of 8.

A possible explanation for more polarization arising when increasing the amount of deletion of extreme users is that the agents that are mainly deleted are the most active agents of the network. Since they keep their activity when being reset they move from moderate opinions to extreme opinions over and over again. While moving through the system they exert influence on the less active nodes and thus

pull them with them towards the extremes. This hypothesis can be tested by again adding penalization by activity to the system as was discussed in section 4.2 with $f(a)$ again being the discussed Heaviside step function. The distributions emerging by combining both manipulations for $\alpha = 1.5$ and $\alpha = 2.5$ can be found in figures 6.3 and 6.4

When now comparing figures 6.1 and 6.3 it is visible that polarization is less strong for the case with activity penalization since there are more nodes between the two peaks for all radii. The PDFs of agent's opinions near the phase-shift in figure 6.4 don't show polarization anymore even when the radius of the discs drawn around the agents is small.

As a last variation of the deletion of users the agents were again deleted as described before but with them now having their activity redrawn from the activity probability distribution. The agents were not penalized by their activity. The resulting PDFs again for $\alpha = 1.5$ and $\alpha = 2.5$ can be seen in figure 6.5. When comparing these PDFs for a radius of 8 with the ones created by the original model but with a maximum activity of 0.1 which can be seen in figure 6.6 it can be observed that the PDFs are very similar.

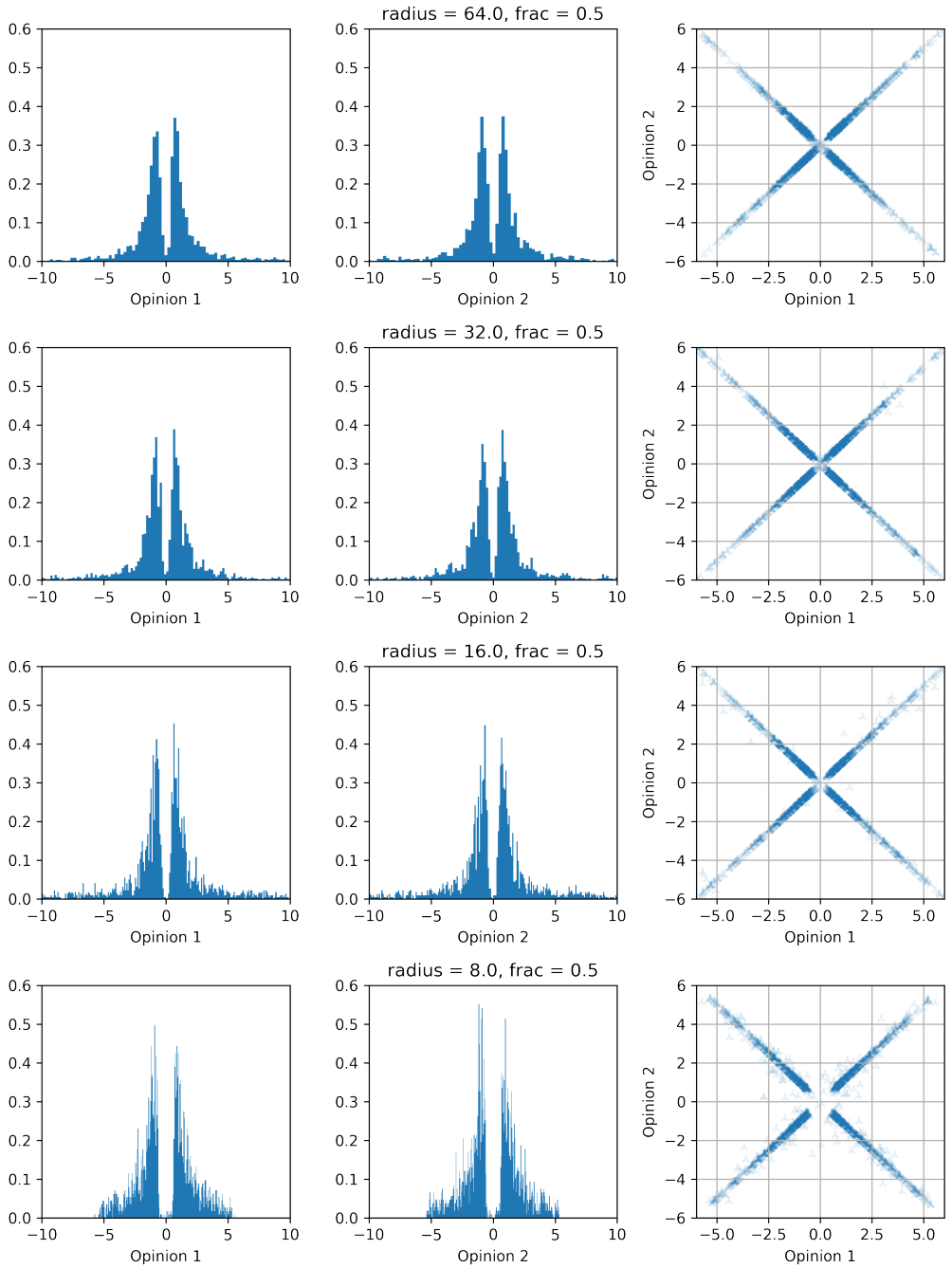

Figure 6.1.: PDFs of agent's opinions for different parameters including deletion of agents and the according opinion-spaces are shown. All simulations were performed for $\alpha = 2.5$ and $\cos(\delta) = 0.0$.

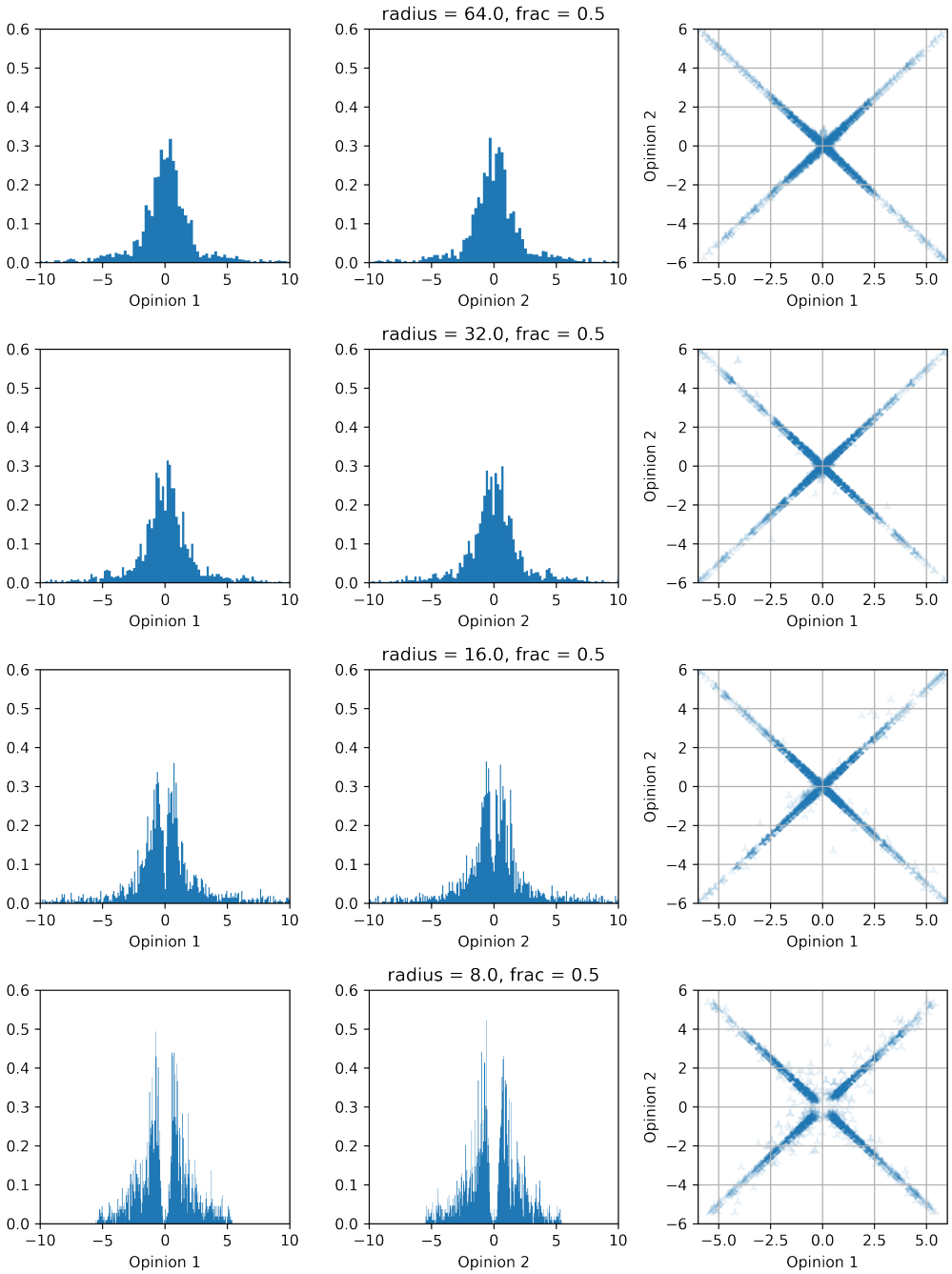

Figure 6.2.: PDFs of agent's opinions for different parameters including deletion of agents and the according opinion-spaces are shown. All simulations were performed for $\alpha = 1.5$ and $\cos(\delta) = 0.0$.

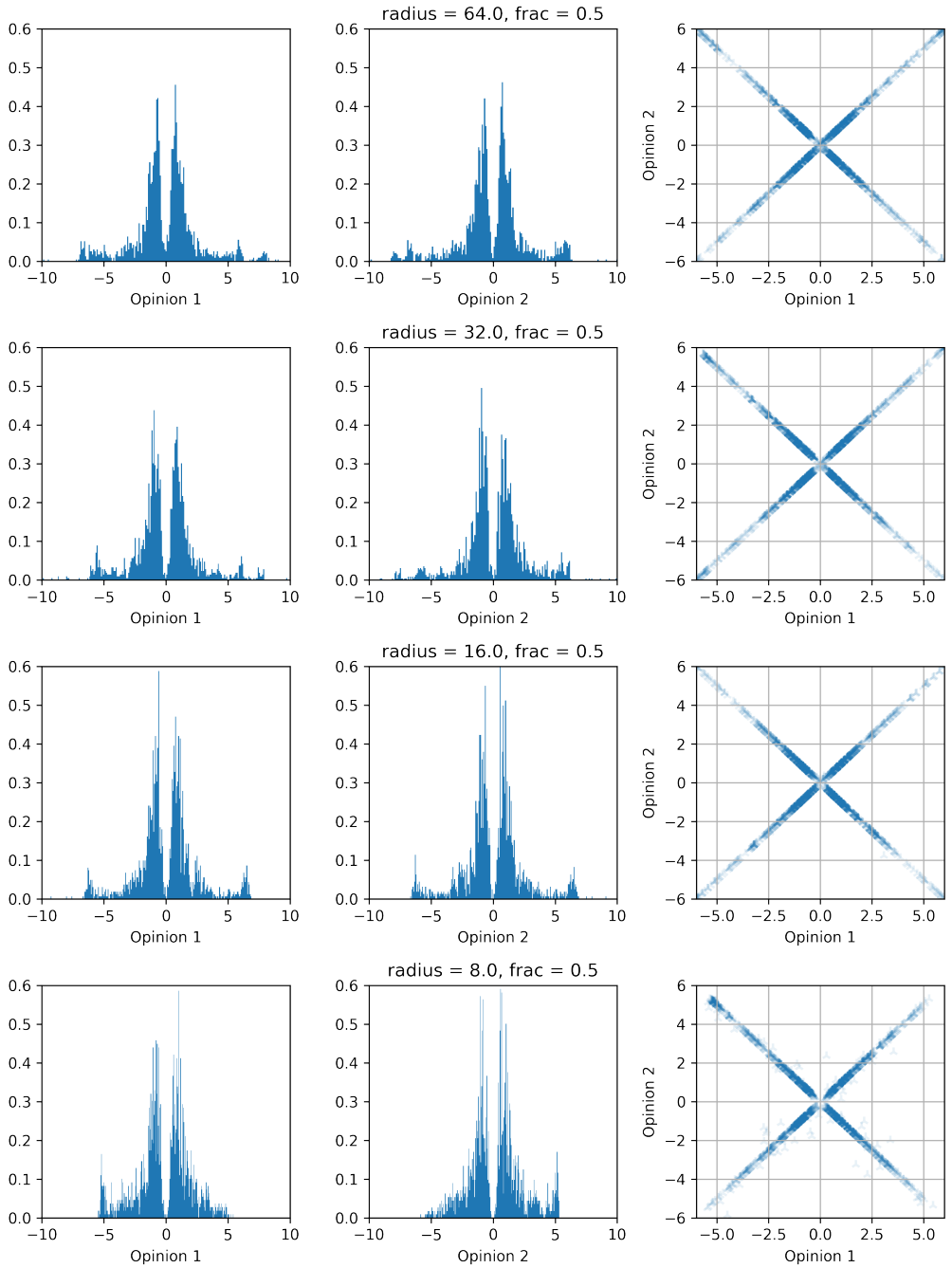

Figure 6.3.: PDFs of agent's opinions for different parameters including deletion of agents and the penalization of very active agents and the according opinion-spaces are shown. All simulations were performed for $\alpha = 2.5$ and $\cos(\delta) = 0.0$.

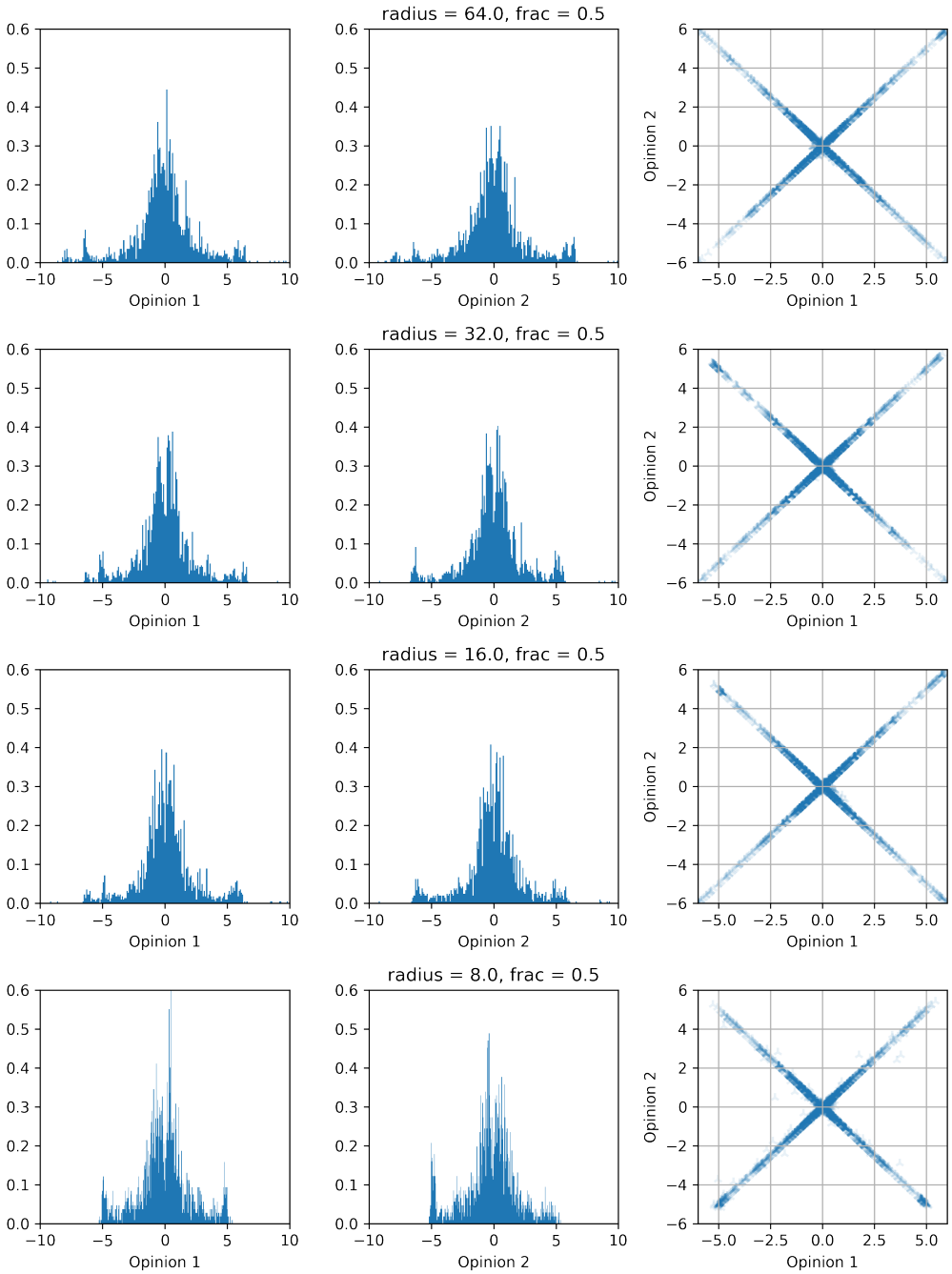

Figure 6.4.: PDFs of agent's opinions for different parameters including deletion of agents and the penalization of very active agents and the according opinion-spaces are shown. All simulations were performed for $\alpha = 1.5$ and $\cos(\delta) = 0.0$.

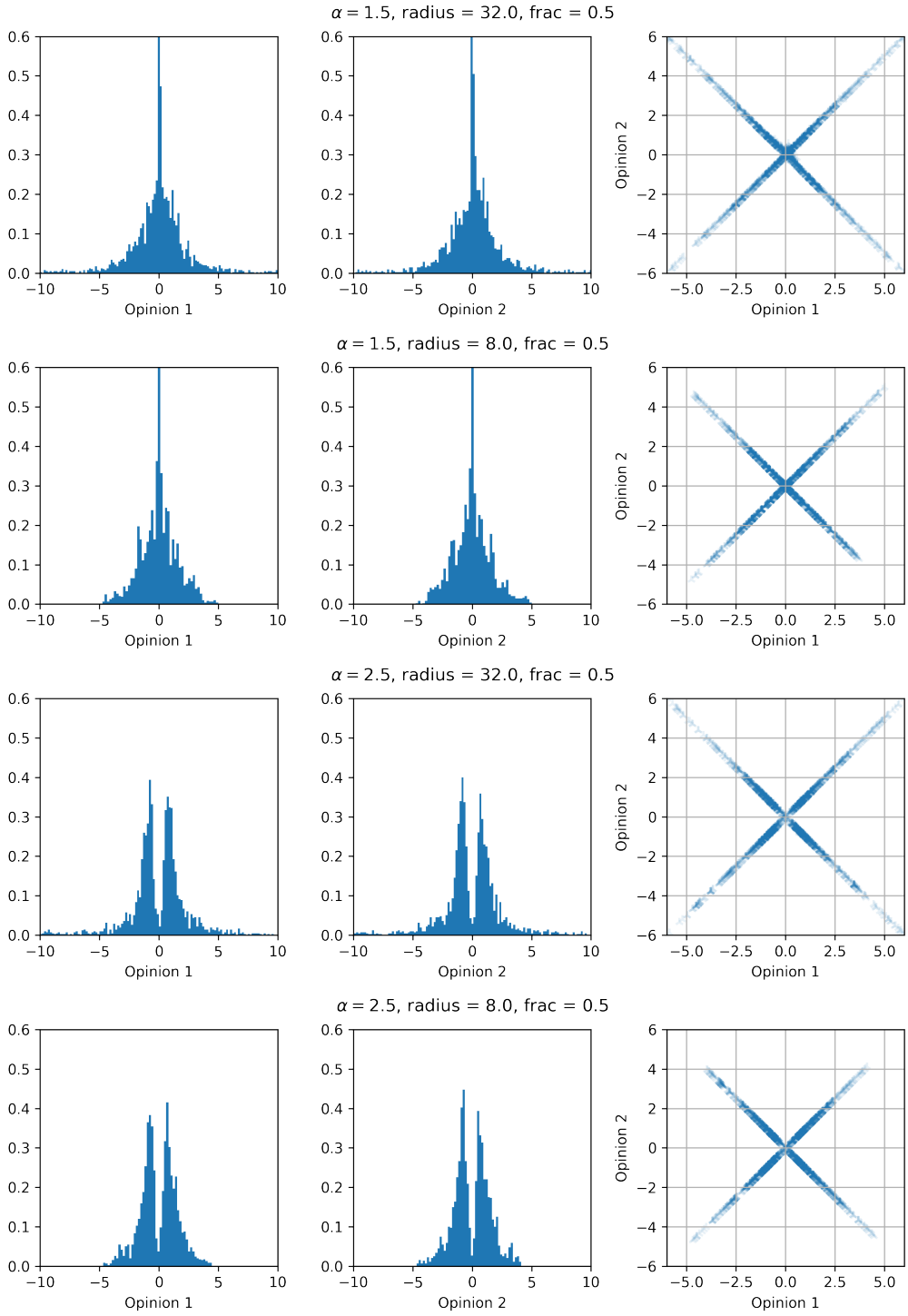

Figure 6.5.: PDFs of agent's opinions for different parameters including deletion of agents and the resetting of their activity and the according opinion-spaces are shown.

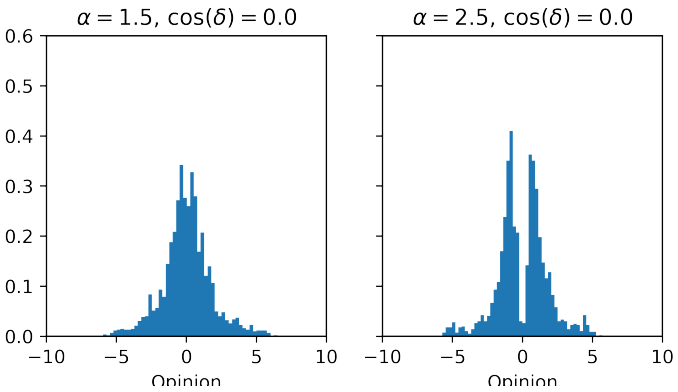

Figure 6.6.: Two PDFs can be seen for $\alpha = 1.5$ and $\alpha = 2.5$. The activity of the underlying network was drawn from the same power-law as before but this time with the power-law being truncated on $[0.01, 0.1]$.

7. Discussion

In the following both the success of the different depolarization methods and the limitations of the model are going to be discussed.

7.1. Success of the Different Methods

7.1.1. Manipulating the Most Active Users

When looking at the results of penalization by activity and by recent connections (sections 4.2 and 4.3) one can see that the phase-space doesn't show any notable changes. Since both methods punish the most active users it can be deduced that very active users don't pull less active users towards the extremes and thus don't matter much for polarization of less active users. It seems to be the case that the most active users don't act as influencers but are instead very influenced by others which can be seen by their opinion being extreme while their influence on the system is not of importance.
The very active users moving towards the extremes makes sense when looking at equation (2.1). Since the active users connect to others much more often than the normal user the social influence term dominates the equation resulting in them having a bigger mean social influence. Because of that they reach a fixed point only when taking on an extreme opinion since the decay term of equation (2.1) is the negative of their opinion.
Their extreme opinion not mattering much for the network results from the capping of the influence of an agent's opinion through tanh. Also since the very active users have an opinion that's often far from the center they interact mostly with other extreme agents which is especially true for the case of strong homophily. If they interact with users that are near the center the distance towards them is so high that users around the center are chosen with roughly the same probability resulting in extreme users also interacting with agents that have a different conviction than

them. This means that the very active users interact with agents that can have all combinations of positive or negative stances towards the topics and not only with the ones who's opinion's signs are the same as the ones of a active user. Thus they pull agents of all convictions towards their opinion but since at the same time extreme agents are present at each extreme their influence on the center users evens out. Because of that the extreme agents don't play a big role in polarizing the lesser active users, which explains why completely negating the influence of users with $a > 0.2$ doesn't make a change to the phase-space.
For the same reasons penalizing recent connections doesn't make a difference. Another reason for the penalization of recent connections not working is that agents have a small reach compared to the network size, $m \ll N$. Thus even when the probability of connecting quickly again is lowered there are still enough other agents close to the recently contacted agents that can now be contacted because of which the penalizing effect is diminished. It is possible that for networks with a higher ratio of m to N the method of penalization of recent connections has a bigger effect. All in all it is clear that simply penalizing the most active users of the network doesn't work when trying to depolarize the network.
When now looking into the effect changing the activity probability distribution to a Gaussian with a mean of 0.5 and a standard deviation of 0.1 has one can see that the phase-space depicts much more polarization and ideology and thus less consensus. This is caused by more agents being more active which leads to them being more prone to social influence. Because of that the social influence term in equation (2.1) is more dominant which is why in turn their fixed point is located at a more extreme opinion since the decay term of the differential equation has to compensate for the stronger social influence. Since now more agents have a bigger absolute opinion the social influence they have on others is still big even for small values of α which leads to polarization happening for controversialness values that previously resulted in consensus.
It is of importance that the number of active users per network iteration was held constant so that both the model with power-law distributed activity and the one with Gaussian distributed activity have the same global activity. If the global activity was bigger for the Gaussian case users would experience even more social influence resulting in even stronger polarisation but this would make the different models incomparable. Since the mean global activity is kept constant the results show that if most users would share the same activity the network would be more

prone to polarization compared to the case of many users not being very active and few exercising most of the networks activity. Thus simply getting rid of the power-law distribution underlying the activity of agents in the network doesn't help depolarizing the network but rather pushes agents further apart into stronger polarization.

7.1.2. Implementing Moderators

When looking at the phase-space created by the model including moderators that actively pull agents closer to the center of the opinion-space one can see that there is a lot of consensus present and much less ideology with polarization being gone completely.

A first explanation for the effect of moderators with a weak pulling force compared to the mean social-influence an agent is prone to is that most of the users experience much less social influence per iteration than the average suggests because of the activity probability distribution being a power-law. Very active agents have a much bigger mean social influence than moderately active agents which reduces the meaningfulness of the comparison of s_m and $\bar{K}$. This means that for most of the agents s_m is a strong influence resulting in the agents having a fixed-point closer to or even directly on the center of opinion-space.

A second explanation could be that multiple moderators were present in the network with all of them being able to attach to the same agents. Because of that some agents could have been pulled much stronger towards the center than others which would lessen the influence they had on others which in turn would have helped depolarizing the network.

This explanation can be rejected by finding that even for having few moderators with a bigger reach the same amount of depolarization is observed. At the same time the explanation can be dismissed for the case of equal picking-probabilities since it is unlikely that many agents are connected to by multiple moderators in each turn which would explain why the observed depolarization is the same for having different amounts of moderators. It simply seems to be the case that the pulling strength is strong enough for most nodes to move towards neutrality with the nodes having low activities being affected the strongest.

The depolarization caused by moderators that used the agent's activities as connection probabilities not being as strong as for the case of equal probabilities can again be explained by the very active agents not having much of an impact on the

less active agents. Thus even though the extreme agents are moderated strongly the less active agents are polarizing each other as they were without the influence of moderators. Because of that the opinion distributions are less fat-tailed but the two peaks that characterize opinion-polarization remain for strong controversialness. Thus consensus is reached for values of $\alpha \approx 2$ since these networks are near the phase-shift to consensus in the original model leading to them ending up as consensus even though the effect of activity-sorting moderators was small on most of the agents of the network.

7.1.3. Moving Influencers

Evaluating the resulting PDFs of agent's opinions from implementing moving influencers in the network it can be deduced that the system wasn't depolarized but instead agents grew more extreme and apart. This can be explained by looking at equation (2.1). In the case of agents i and j connecting i is influenced in a way that his opinions grow towards the conviction of agent j. The pull towards the conviction of agent j doesn't work in an assimilative [11] way, meaning that agent i and j move closer together, but instead in the opinion vector of agent i moving closer to the extremes of the conviction agent j has. Because of this agent i grows more extreme even if his opinions are similar to or even more extreme than the opinions of agent j.

Thus when now an influencer that sits at a moderate opinion connects repeatedly to agents that are near him he causes these moderate agents to grow more extreme because of his influence creating a pull towards the extremes of the influencer's conviction. This is why generally the implementation of the moving influencers causes polarization instead of depolarization.

What is left to be explained is that for less influencers with more reach less polarization arises compared to the case of having more influencers with less reach while $N_m \cdot m_{mod}$ remains constant between the compared distributions.

The behavior can be explained by homophily. If many influencers with less reach sit at the same point in opinion-space they are likely to attach to similar agents that are placed closely to them. Through this they cause a stronger push away from neutrality for the agents around them compared to the case of few influencers having bigger reach. Especially when the many influencers are sitting close to the center of opinion-space they push away the agents that would otherwise sit close to the center while the agents that originally have had an extreme opinion remain at their

position. Thus the single peak that was originally present for the case of consensus is split into two extremes which is what can be observed in the final distributions.

7.1.4. Removing Extreme Agents From the Network

The effect of the removal of nodes with an extreme opinion from the network was looked into by the changes that the removal made to the PDFs of agent's opinions for both a polarized and a consensus like state. The stricter the conditions for staying in the network became the more polarized both the originally polarized and the originally consensus-like PDFs got. This is unintuitive since removing extreme nodes and initializing new agents with moderate opinions should supposedly help with depolarizing a network. The behavior can be explained by the very active agents which are most prone to deletion moving through the system multiple times instead of moving to the extremes and staying there.
When staying at the extremes the very active agents don't influence the network much anymore since they are so far from others that they rarely interact with moderate agents because of homophily. In contrast when an active agent is reset and moves to the extremes again the agent is close to moderate agents when being reset which results in him exerting influence on the moderate agents. Thus the moderate agents are influenced more strongly over multiple iterations which leads to a more extreme opinion of the agents. The hypothesis was tested by removing the influence highly active agents had on the network which lead to no further polarization which indicates that the hypothesis is true.
Implementing the redrawing of an agent's activity if he is reset doesn't depolarize the network but rather clears it from the extreme agents. This is again due to the agent's opinion being correlated to his activity through the mean social-influence he is exposed. Thus an agent is reset as long as his activity causes his opinions to grow too distant from the other agents and stabilizes only when having an activity that places him close to the other agents. Because of that the network reaches a stable state when all highly active agents have taken on a lower activity leading to the opinion distribution being unchanged around the center of opinion-space with the tails being cleared.
Furthermore since the activity of reset agents is drawn from the same power-law distribution as before the stable network is close to a one with that has an underlying activity probability distribution that is drawn from $[\epsilon_1, \epsilon_2]$ with $\epsilon_1 < \epsilon_2 < 1$ where ϵ_2 depends on the strength of the criterion for deletion.

When now evaluating the results it can be concluded that deleting extreme agents from the network doesn't help depolarizing it, especially if the deletion is implemented without penalizing the very active agents of the network.

7.2. Limitations of the Model

Now the limitations of the original model by Baumann et. al. have to be discussed so that the application of the results to real systems can be discussed. Three main points can be found.
The first critic is that the fixed points of nodes are set depending on the activity of the agent. This can be deduced from equation (2.1) and figure 5.1. Since the exponent of the power-law depicted in the figure has the same exponent of the activity distribution it can be theorized that the activity of an agent roughly determines the mean social influence an agent receives.
This hypothesis is backed up by figure 4.1 since the more active agents are the more extreme their opinion is. This again makes sense when looking at equation (2.1). A strong social influence results in a strong opinion since the decay term consisting of the opinion of an agent has to compensate his social influence in order to reach stability. Thus the absolute of an agent's opinion, his activity and the mean-social influence he is exposed to are correlated.
This now makes it hard to influence agents by implementing moderators or influencers as was done in section 5.2 since an influencer can move an agent but in order for the agent to stray from his fixed point the agent has to be periodically influenced. Without periodical disturbances the agent will return to his pre-determined absolute of opinion, making the model very stable against disturbances. This makes the model unrealistic since for example in reality people can move from extreme opinions to moderate opinions and vice versa and gravitate back to their new opinions instead of their old ones.
The second point is that agents don't move closer to each other but rather influence others towards their opinions. This is a problem for the efficiency of influencers as they were implemented in section 5.2. Even if influencers sit at moderate opinions they polarize because they don't pull agents towards them but rather pull them towards the extreme of the quadrant in opinion-space they are sitting in meaning that even moderate agents that already sit close to an influencer are polarized more strongly since they receive more social influence from the influencers.

Because of that changing the activity distribution to a Gaussian distribution or deleting extreme users from the network and resetting them doesn't result in less polarization but rather in stronger polarization.
The third property that is to be criticized is that the reach agents have is constant instead of being power-law distributed which is what can be found in social systems [13]. Because of that the very active agents that should be acting as influencers are strongly influenced instead of being strongly influential. This effect is increased by the opinions $x_i^{(\nu)}$ not being truncated to stay in a certain interval and thus taking on very extreme values combined with strong homophily being present in the network. With what was found before this means that the very active agents move into corners and mostly influence other very active agents and thus essentially leave the moderate agents alone. Because of that it is of question to which social systems and especially to which social media [2] the model and especially the results found before are applicable to.
Since the model is unable to capture the effect hubs have on a social system the results can't be applied to platforms where power-law distributed reach is present and of importance, for example Twitter or Instagram. Some of the results could probably be applied to social media structures like Reddit where users don't follow users but rather communities [7]. Thus their reach is possibly not power-law distributed and the activity of each user could be of more importance. At the same time users on Reddit are not only influenced by other users but also by posts which could potentially create new dynamics.
What can be deduced from the results is that moderators that effectively move agent's opinions towards a center are going to be helpful in depolarizing social systems but it is unclear how to that in reality. For big social media platforms like Facebook and Instagram a moderator per 100 users would mean having to have millions of moderators which is unrealistic. Possibly moderators can be used more efficiently in networks were most people are influenced by few hubs by especially moderating these hubs which was the idea behind the moving influencer of section 5.2.
The other method that seems to be probable for depolarizing a system is reducing the activity users spend on social networks and especially on social media. If people were less exposed to polarizing opinions they could possibly stay more moderate and thus be more open towards other's opinions which could in turn make consensus more probable.

This approach is unlikely to be implemented by the companies that run social media as long as the company makes money of people being online and active. The biggest source of income for companies running the biggest social media networks like Instagram, Facebook, Reddit and Youtube is advertising [1, 8, 12, 14, 18, 25] causing the companies to get more revenue when people spend more time on their social media platform. Thus the approach could be implemented by the users of social media. They could sacrifice their time spent on social media in order to avoid opinion polarization but since social media plays an increasingly large role in the life of people that is also unlikely.

8. Future Research

In this chapter possible future research that should be done on the results found in this thesis and on mechanisms for depolarizing is proposed.
The first and most important thing is evaluating the effect the proposed methods have on models with different structures. As mentioned in the former chapter very active agents don't influence most of the other agents. Thus the proposed methods should be tested on models in which influencers do have a big impact on the whole network. The methods should also be tested on models with assimilative influence and on ones with similarity bias [11] as in these models moderators would probably be more effective since they would actually draw users towards them instead of influencing them into a certain direction.
It would also be interesting to see the effect some of the methods have on models that are based on rather static networks that can implement information cascades [6] like the model proposed by Prasetya and Murata [24]. This would test the effect of methods of depolarization on social media structures like Twitter [27].
Other things that could lead to more insight would be making further changes to the model by Baumann et. al. It could for example be relevant to change the behavior of the agents of the network in a sense that user's opinions don't decay back to a neutral opinion but rather to a certain opinion x_i^s. This opinion could at first be an agents initial opinion but x_i^s should be able to be changed by social influence. Thus moderators could effectively change the opinion that agents decay back to because of which influencers as described in section 5.2 would potentially make big changes to the system.
Another modification that could be done to the model is implementing a power-law distribution not only to the agent's activity but also to their reach in order to possibly create more influential agents that reach most of the network despite being very active.
A last modification that could be of interest would be to make the model directed [22, p.24] meaning that agents that are activated only influence the agents that

they are connecting to while not being influenced by these connected agents. Thus influencers would likely not grow as extreme as they do in the original model which would cause them to play a bigger role in the formation of polarization.

A. Effects of Network Size, Stability Analysis and Choice of Parameters

Looking at figure A.1 one can see that for constant m the position of the peaks remains unchanged for a varying network size. For larger numbers of N the peaks become more elaborate but the shape of the distribution remains the same. For $m = N/1000$ the peaks grow further apart. Since the simulations are computationally heavy and since the shapes of the peaks are elaborate enough $N = 2500$ was chosen as a good compromise of the network size for the simulations. Also $m = 10$ was chosen so that the peaks always formed approximately around an absolute opinion of 1.

Analyzing the effect different β with $\beta \ll 5$ have on the PDFs of agent's opinions it can be seen in figure A.2 that there are no notable differences in the PDFs for $\beta > 5$. The shape of the PDFs remains the same for depicted β. Thus $\beta = 5$ was chosen for all simulations.

In figure A.3 the distance between each agent's mean opinion vector of 10 iterations and his mean opinion vector of the former 10 iterations are shown for 1000 iterations. It can be observed that the networks stabilized after 500 iterations which can be deduced when looking at the mean distance between agent's opinion vectors which decays to a constant value that is close to zero. The agent's distance between mean vectors is always fluctuating especially for very active agents since the agents opinion is changing in each iteration. When agents aren't connected to others they move towards the center of opinion-space but move to the extremes as soon as they are prone to social influence which is why each agent fluctuates around his fixed point. In order to ensure stability for parameters that cause agents to be polarized more strongly, meaning $a \gg 1$ and $\cos(\delta) \gg 0$, all simulations were run for at least 1000 iterations.

With the results of the foregoing parameter analysis the parameters that were chosen for all simulations in the thesis are $N = 2500$, $m = 10$ and $\beta = 5.0$ as well as $\gamma = 2.1$, $\epsilon_1 = 0.01$, $K = 3$ and $T = 2$.

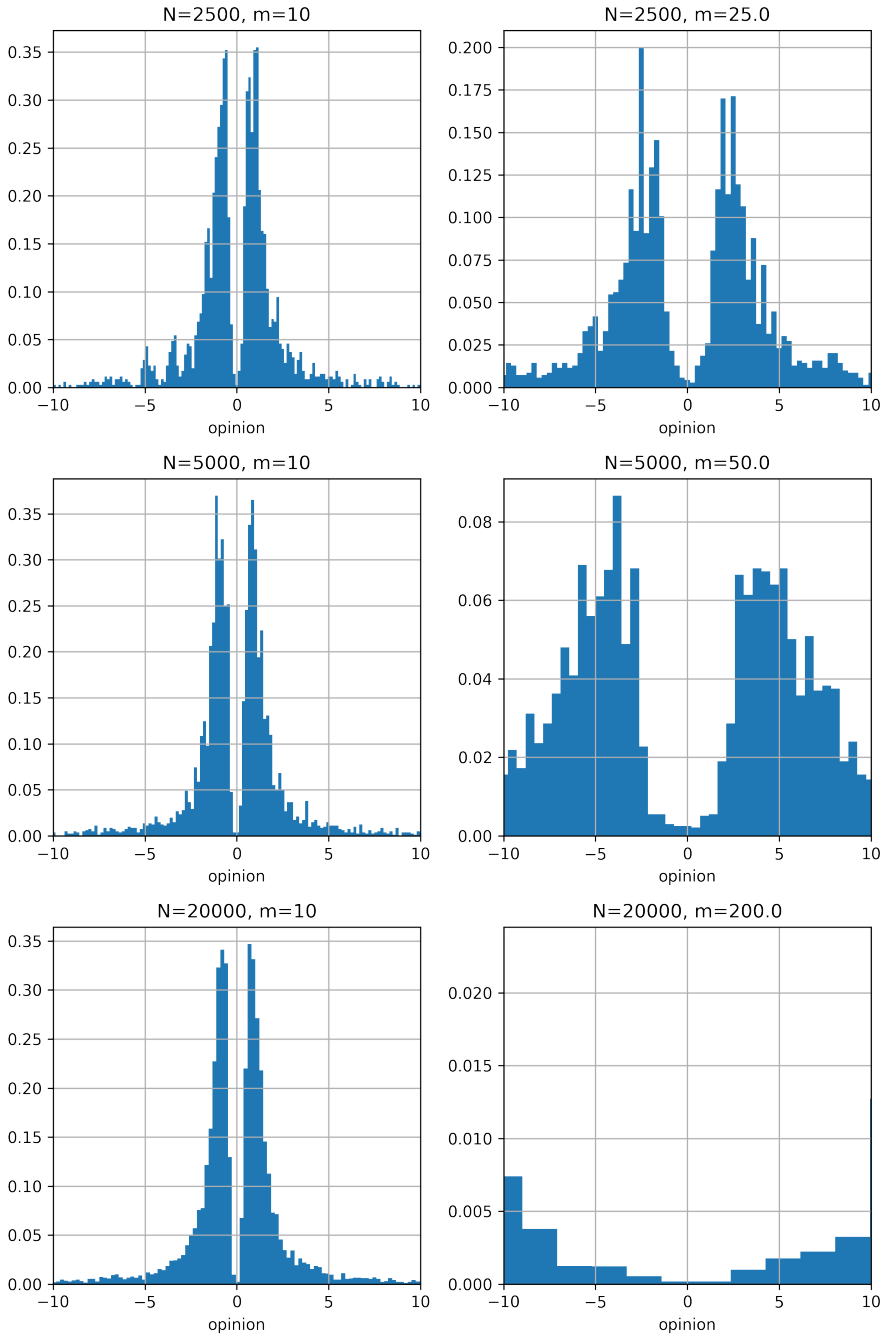

Figure A.1.: The PDFs of agent's opinions are shown for simulations with $\alpha = 3.0$ and $\beta = 5.0$ for different N and m.

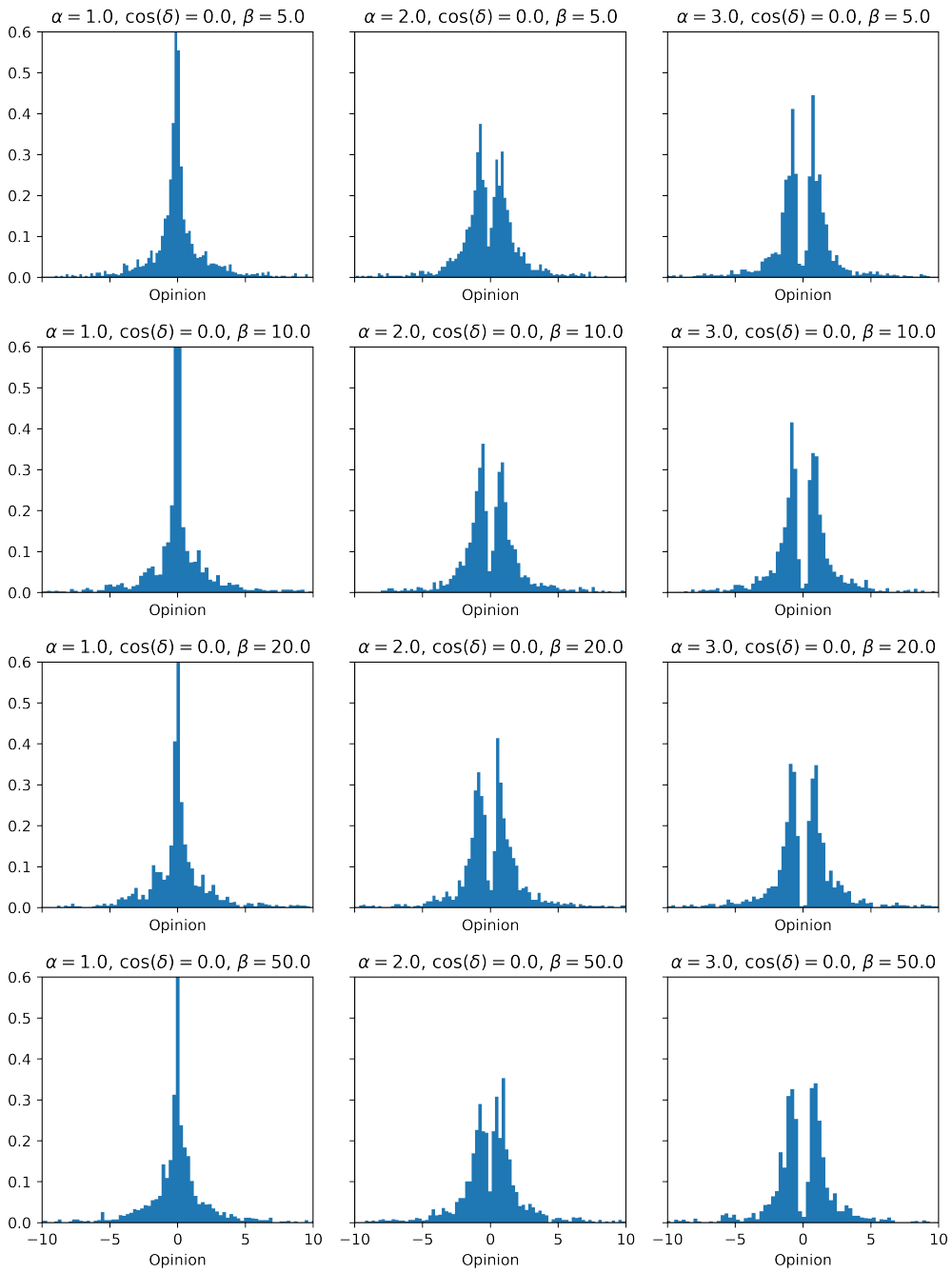

Figure A.2.: Different PDFs of agent's opinions are shown for varying parameter tuples $(\alpha, \cos(\delta), \beta)$.

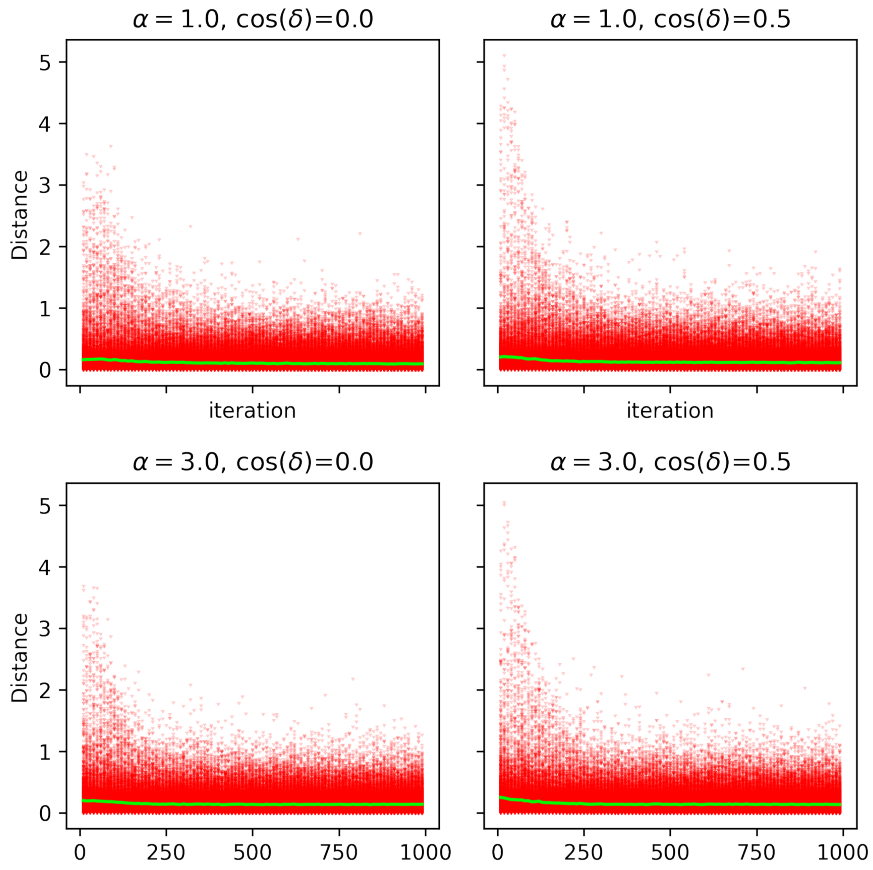

Figure A.3.: The distance of each agent between his mean opinion vector of 10 iterations and his mean opinion vector of the former 10 iterations can be seen for different parameter combinations. The mean distance of all agents is plotted in green.

B. Agglomerative Classifier

At the top of figure B.1 an example of the final nodes that include at least 200 agents produced by the agglomerative algorithm for a polarized state are presented. The zones that are used to determine where a final node lies in the opinion space are colored in. The presented example is classified as polarization, since at least one final node is present in each of the four zones and since none of the final nodes are within the zone for consensus, which can be seen at the bottom of of figure B.1. The example on the bottom is one that is close to transitioning to consensus which is why final nodes emerge not only in the four corners but also in the middle of opinion space. Since a final node is found in the zone around the center it is counted whether the nodes outside of the zone include more agents than the node in the middle. Since this is the case for this example the distribution is classified as polarized. If more agents would be included in the center final node the distribution would be classified as consensus.

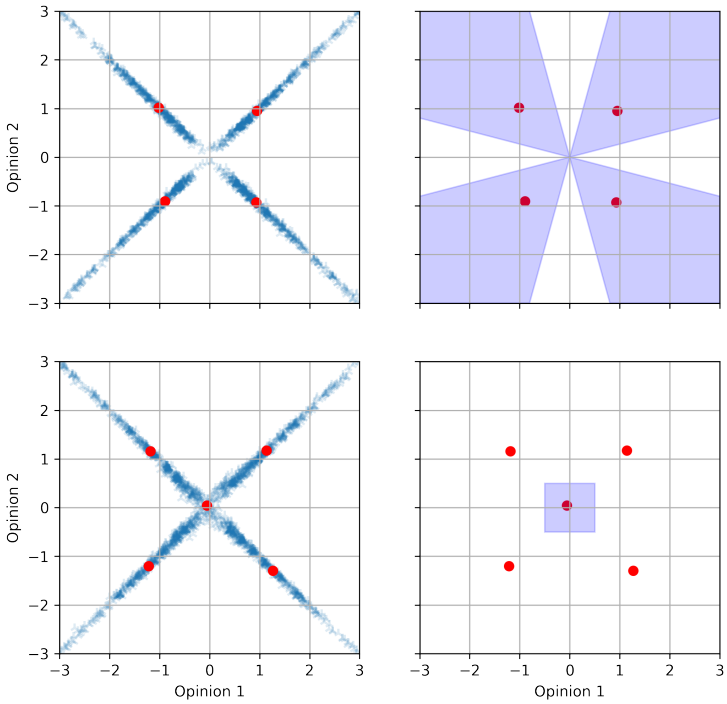

Figure B.1.: The top two figures show the final nodes found by the agglomerative algorithm for a polarized state ($\alpha = 3.0$). On the left the original agent distribution with the final nodes in red is shown and on the right the final nodes with the zones used for classifying in which area of the opinion-space they are in can be found. The bottom two figures show the final nodes for a state with $\alpha = 1.8$. On the left again the agent's original distribution with the final nodes found by the agglomerative algorithm in red can be seen while on the right only the final nodes and the zone in blue used for classifying whether nodes contribute to a classification of consensus state or not are depicted.

Bibliography

[1] How youtube makes money. https://www.youtube.com/howyoutubeworks/our-commitments/sharing-revenue/. Accessed: 2022-10-11.

[2] Thomas Aichner, Matthias Grünfelder, Oswin Maurer, and Deni Jegeni. Twenty-five years of social media: A review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, Behavior, and Social Networking*, 24(4):215–222, April 2021. doi: 10.1089/cyber.2020.0134. URL https://doi.org/10.1089/cyber.2020.0134.

[3] Katja Albada, Nina Hansen, and Sabine Otten. Polarization in attitudes towards refugees and migrants in the netherlands. *European Journal of Social Psychology*, 51(3):627–643, April 2021. doi: 10.1002/ejsp.2766. URL https://doi.org/10.1002/ejsp.2766.

[4] Geeta Arora, Varun Joshi, and Isa Sani Garki. Developments in runge–kutta method to solve ordinary differential equations. In *Recent Advances in Mathematics for Engineering*, pages 193–202. CRC Press, March 2020. doi: 10.1201/9780429200304-9. URL https://doi.org/10.1201/9780429200304-9.

[5] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. Emergence of polarized ideological opinions in multidimensional topic spaces. *Physical Review X*, 11(1), January 2021. doi: 10.1103/physrevx.11.011012. URL https://doi.org/10.1103/physrevx.11.011012.

[6] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, October 1992. doi: 10.1086/261849. URL https://doi.org/10.1086/261849.

[7] Francesco Cauteruccio, Enrico Corradini, Giorgio Terracina, Domenico Ursino, and Luca Virgili. Investigating reddit to detect subreddit and author stereotypes and to evaluate author assortativity, 01 2021.

[8] Arieez Dutta. Youtube business model | how does youtube make money? https://www.feedough.com/youtube-business-model-how-does-youtube-make-money/, June 2022. Accessed: 2022-10-11.

[9] Paul Fieguth. *An Introduction to Complex Systems*. Springer International Publishing, 2017. doi: 10.1007/978-3-319-44606-6. URL https://doi.org/10.1007/978-3-319-44606-6.

[10] Helmut Fischer and Helmut Kaul. *Mathematik für Physiker Band 1*. Springer Berlin Heidelberg, 2018. doi: 10.1007/978-3-662-56561-2. URL https://doi.org/10.1007/978-3-662-56561-2.

[11] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4), 2017. doi: 10.18564/jasss.3521. URL https://doi.org/10.18564/jasss.3521.

[12] Kamil Franek. How facebook makes money: Business model explained. https://www.kamilfranek.com/how-facebook-makes-money-business-model-explained/, April 2021. Accessed: 2022-10-11.

[13] Piotr Fronczak. Scale-free nature of social networks. In *Encyclopedia of Social Network Analysis and Mining*, pages 2300–2309. Springer New York, 2018. doi: 10.1007/978-1-4939-7131-2_248. URL https://doi.org/10.1007/978-1-4939-7131-2_248.

[14] Anisha Gera. How does reddit make money? business model breakdown. https://thestrategystory.com/2021/09/23/how-does-reddit-make-money-business-model/, September 2021. Accessed: 2022-10-11.

[15] Michael Hout, Stuart Perrett, and Sarah K. Cowan. Stasis and sorting of americans' abortion opinions: Political polarization added to religious and other differences. *Socius: Sociological Research for a Dynamic World*, 8: 237802312211176, January 2022. doi: 10.1177/23780231221117648. URL https://doi.org/10.1177/23780231221117648.

[16] Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 22(1):129–146, May 2019. doi: 10.1146/annurev-polisci-051117-073034. URL https://doi.org/10.1146/annurev-polisci-051117-073034.

[17] Sebastian Jungkunz. Political polarization during the COVID-19 pandemic. *Frontiers in Political Science*, 3, March 2021. doi: 10.3389/fpos.2021.622512. URL https://doi.org/10.3389/fpos.2021.622512.

[18] Joe Keeley. How do social networks make money? explained. https://www.makeuseof.com/tag/how-do-social-networks-make-money-case-wondering/, April 2022. Accessed: 2022-10-11.

[19] Namkje Koudenburg, Henk A. L. Kiers, and Yoshihisa Kashima. A new opinion polarization index developed by integrating expert judgments. *Frontiers in Psychology*, 12, October 2021. doi: 10.3389/fpsyg.2021.738258. URL https://doi.org/10.3389/fpsyg.2021.738258.

[20] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: A comparative analysis. *Physical Review E*, 80(5), November 2009. doi: 10.1103/physreve.80.056117. URL https://doi.org/10.1103/physreve.80.056117.

[21] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, August 2001. doi: 10.1146/annurev.soc.27.1.415. URL https://doi.org/10.1146/annurev.soc.27.1.415.

[22] Filippo Menczer, Santo Fortunato, and Clayton A. Davis. *A First Course in Network Science*. Cambridge University Press, January 2020. doi: 10.1017/9781108653947. URL https://doi.org/10.1017/9781108653947.

[23] MEJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351, September 2005. doi: 10.1080/00107510500052444. URL https://doi.org/10.1080/00107510500052444.

[24] Hafizh A. Prasetya and Tsuyoshi Murata. A model of opinion and propagation structure polarization in social media. *Computational Social Networks*, 7(1), January 2020. doi: 10.1186/s40649-019-0076-z. URL `https://doi.org/10.1186/s40649-019-0076-z`.

[25] Muaaz Qadri. Instagram business model case study. `https://whatisthebusinessmodelof.com/business-models/how-instagram-makes-money/`, June 2021. Accessed: 2022-10-11.

[26] Saul Stahl. The evolution of the normal distribution. *Mathematics Magazine*, 79(2):96–113, April 2006. doi: 10.1080/0025570x.2006.11953386. URL `https://doi.org/10.1080/0025570x.2006.11953386`.

[27] Io Taxidou and Peter M. Fischer. Online analysis of information diffusion in twitter. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, April 2014. doi: 10.1145/2567948.2580050. URL `https://doi.org/10.1145/2567948.2580050`.

Thanksgiving

Here I want to express my gratitude towards all who helped me write my bachelor thesis.
At first I have to thank Dr. Joao Pinheiro Neto and Riccardo Carlucci for supervising me while working on my thesis. Our discussions about ideas and problems that arose in the process and their guidance as well as their constructive criticism were a huge help.
I also want to thank Dr. Knut Heidemann for letting me participate in his research group and for guiding me towards my supervisors.
My gratitude goes to my friend and fellow student Vincent Brockers for our helpful exchange of ideas and for him proofreading the thesis.
At last I want to thank my parents for always supporting me on my way. Without them I would not have gotten so far.

Erklärung Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder sinngemäß aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen wurden von mir beachtet. Eine gegebenenfalls eingereichte digitale Version stimmt mit der schriftlichen Fassung überein. Mir ist bewusst, dass bei Verstoß gegen diese Grundsätze die Prüfung mit nicht bestanden bewertet wird.

Göttingen, den 25. Oktober 2022

(Robin Cedric Danek)