POLITECNICO DI TORINO

MASTER'S DEGREE IN PHYSICS OF COMPLEX SYSTEMS

Modeling Multidimensional Political Ideologies on Reddit

Author:
Ernesto A. P. COLACRAI

Supervisors:
Dr. Gianmarco DE FRANCISCI MORALES
Dr. Michele STARNINI

Master's Thesis realized in collaboration with

CENTAI Institute

POLITECNICO DI TORINO

Abstract

Master's Degree in Physics of Complex Systems

Modeling Multidimensional Political Ideologies on Reddit

by Ernesto A. P. COLACRAI

The prevalent perspective in quantitative research on opinion dynamics flattens the landscape of the online political discourse into a traditional left–right dichotomy. While this approach helps simplify the analysis and modeling effort, it also neglects the intrinsic multidimensional richness of ideologies. In this study, we analyze social interactions on Reddit, a social news aggregator and forum platform, under the lens of a multi-dimensional ideological framework: the political compass. The Reddit chosen communities for the present study are `/r/PoliticalCompass` and `/r/PoliticalCompassMemes`, two communities dedicated respectively to discussing political views through two-axis ideologies and humorous commentary on such ideologies. Given a rich dataset of over 8 million comments posted during 2020–2022, we aim to investigate the way interactions on social media align with a fine-grained characterization of users in terms of their ideological coordinates and demographic features. By leveraging their self-declarations, we disentangle the ideological dimensions of users into economic (left–right) and social (libertarian–authoritarian) axes. In addition, we characterize users by their demographic attributes (age, gender, and affluence). After reconstructing the interaction network for both communities, we find significant homophily for interactions along the social axis of the political compass and demographic attributes. Compared to a null model, interactions among individuals of similar ideology surpass expectations by 6%. In contrast, we uncover a significant heterophily along the economic axis: left/right interactions exceed expectations by 10%. Furthermore, heterophilic interactions are characterized by a higher language toxicity than homophilic interactions, which hints at a conflictual discourse between every opposite ideology. Our results help reconcile apparent contradictions in recent literature, which found a superposition of homophilic and heterophilic interactions in online political discussions. By disentangling such interactions into the economic and social axes we pave the way for a deeper understanding of opinion dynamics on social media.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Collective Phenomena and the Role of Social Media

Our world is filled with situations where many interacting entities exhibit coordinated, large-scale behavior emerging from their interactions. Those behaviors cannot be simply explained by studying the individual components alone but rather arise from the complex interplay between them. Collective phenomena find roots in statistical physics, but expand to other fields, from the swirling galaxies in the cosmos to the mesmerizing dance of a starling flock [60, 5], pervading the natural world and even extending to the intricate dynamics of human societies and animal groups.

Social media platforms and online communities have become integral parts of our lives, generating massive amounts of data that reflect our interactions, opinions, and behaviors. These platforms act as virtual laboratories for computational social scientists, providing unique opportunities. By studying collective phenomena we gain valuable tools and perspectives for exploring social phenomena in the complex and dynamic world of social media and online interactions. This is the main focus of computational social science, an interdisciplinary approach that holds immense potential for advancing our understanding of human behavior in the digital age, enabling us to address pressing social challenges and build a better future. It employs computational methods to analyze massive datasets generated from human activities, particularly those captured through online platforms. By studying these "digital footprints," researchers gain insights into large-scale social phenomena like the formation of public opinion, the spread of information and misinformation, the emergence of social movements, and the evolution of social networks. Understanding the dynamics of opinion formation in a population is a goal shared by researchers from different disciplines, from social and computer science to physics.

1.2 Background and Related Work

Numerous studies have revealed the presence of opinion polarization in political discussions [45]: the phenomenon whereby two distinct groups tend to have opposite and potentially extreme views on a specific controversial topic [46], spanning religion [66], race, climate change, political ideology [40, 31], and more [56].

When social interactions among individuals are taken into account, we often observe value homophily: individuals prefer to interact with peers that hold similar opinions [16, 31, 17]. The combination of opinion polarization and homophilic interactions leads to *echo chambers* [34, 15], a situation where existing beliefs can be reinforced by exposure to similar opinions. Echo chambers, in turn, contribute to opinion polarization by reinforcing ideological separation and strengthening the social identity of opposing groups [35, 47, 22]. These phenomena are easy to observe on social media such as Facebook or Twitter, where people share their opinions in more informal settings [16, 32, 2, 12, 31].

Other studies, however, find contrasting results [25, 36, 3, 62]. For instance, researchers have observed heterophilic interactions between Clinton and Trump supporters on Reddit, a preference for cross-cutting political interactions that contradicts the echo chamber narrative [20]. Likewise, some scholars attribute opinion polarization to demographic and socioeconomic factors rather than to social media [26, 59, 49]. Indeed, the profound disparities and imbalances found in our society, such as by gender, race, age, and affluence, have become more closely tied to party affiliation and ideological stances [13, 53, 58]. This phenomenon, known as partisan sorting [44], may also be one of the causes of political polarization [62].

The answer to this apparent contradiction may lie in recognizing the intrinsic multidimensional nature of opinion dynamics. Indeed, the process of opinion formation builds upon views on multiple topics, as they might be discussed at the same time. When considering multiple topics, an interesting phenomenon frequently observed is issue alignment, i.e., the presence of correlations between opinions on different topics, particularly along the left–right dimension [29, 27]. For instance, individuals with strong religious beliefs tend to oppose abortion legalization [1], and various other non-trivial correlations manifest [4, 11, 24]. However, the prevalent view in quantitative research on opinion dynamics has focused on the simplest case of one-dimensional opinions concerning a single topic, both at the level of the analysis [6, 16] and modeling efforts [21, 37, 18, 9]. Only recently researchers have started taking into account a comprehensive multidimensional modeling framework for opinion dynamics [8, 57, 14].

1.2.1 The Political Compass Framework

Many systems can be used to characterize people's political ideologies. One of the most common political spectrum frameworks is the political compass[1] [43]. This scheme includes two ideological axes: one for economic values and resource allocation (left–right) [38], and one for social values and personal freedom (libertarian–authoritarian).

The **economic** or *distributive* **axis** measures possible opinions of how people should be endowed with resources. The *left* (equality) pole is defined as the view that assets should be redistributed by a cooperative collective agency: the state in the socialist tradition, or a network of communes in the libertarian or anarchist tradition. The *right* (liberty) pole is defined as the view that the economy should be left to the market system, to voluntarily competing individuals and organizations. This is the classical left–right conflict that dominated the Cold War [43, 41, 42, 52].

The **social axis**—cross-cutting the first one—is concerned with values of fraternity, understood as axiological principles driving institutionalization, community, forms and actors of democracy, and the quality of the process of collective outcomes. This dimension measures possible political opinions either in a communitarian or procedural sense, considering the appropriate amount of personal freedom and participation: *libertarianism* is defined as the idea that personal freedom as well as voluntary and equal participation should be maximized. This would entail the full realization of liberty in a democratic sense. Parts of that view are ideas like autonomous, direct democratic institutions beyond the state and market, the transformation of gender roles, and self-determination over traditional and religious orders. On the opposite end of the axis, *authoritarianism* is defined as the belief that authority and religious or secular traditions should be complied with. Equal participation and a free choice of personal behavior are rejected as being against human nature, or against necessary hierarchies for a stable society [43, 41, 42, 52].

[1] https://politicalcompass.org/

1.3 Research Motivation and Objectives

Existing research shows two contradictory scenarios. On one hand, homophily with polarization leads to echo chambers. On the other hand, heterophilic interactions are observed in political discourse on online social media, seemingly contradicting the first case. The seemingly contradictory findings of homophily and heterophilic interactions can be reconciled by considering the multifaceted nature of ideologies. People may interact with others who share similar views on some issues (homophily) while holding different views on others (heterophily).
This thesis aims to model user interactions on Reddit, disentangling them along two ideological axes of the political compass, as described in Section 1.2.1.

Given these considerations, our research investigates the alignment between user interactions on social media and their features. We focus on two key aspects of user features:

- **RQ1:** How do interactions on social media align with a fine-grained characterization of users in terms of their ideological positions?
- **RQ2:** How do interactions on social media align with the characterization of users in terms of their demographic features?

Furthermore, we delve deeper into the relationship between inferred 'partisanship' from demographics and self-declared economic ideologies:

- **RQ3:** How does the 'partisanship' dimension inferred from demographic features align with the economic ideologies obtained from user self-declarations?

Finally, we explore the underlying drivers of interactions, potential confounding factors, and the conflictual level of interactions:

- **RQ4:** What are the key drivers of social network interactions? Additionally, how do Reddit-specific characteristics (confounders) play a role in these interactions?
- **RQ5:** How conflictual are interactions between and within ideological groups?

The work is organized as follows:

i We examine a large dataset of political interactions on Reddit, one of the most popular social media platforms (Section 2.1), where users provide self-declared ideological positions on the political compass (Section 2.2), and characterize it both in terms of ideological and demographic composition. The use of self-declaration allows us to avoid inference techniques and achieve greater accuracy for the identification of the ideology of users. We infer instead socio-demographic features (age, gender, affluence, and partisanship) through an unsupervised learning approach (Section 3.2).

ii Then, we examine the interactions among users along the two ideological axes (Section 4.2.2) and according to their demographics (Section 5.2.1). Comparing them to those obtained from a null model of interactions (Section 4.3), our empirical observations show marked homophily on the social axis (Section 5.1.1) and for demographic characteristics (Section 5.2.2). Conversely, the interactions are more heterophilic than expected on the economic axis (Section 5.1.2).

iii Finally, by analyzing the toxicity of the language associated with the interactions, we find that ideological cross-group interactions present higher-than-expected toxicity (Section 5.4). Within-group interactions, on the contrary, show toxicity levels lower than the one expected from the null model, hinting at a certain degree of social affinity [64].

Overall, this multidimensional approach allows us to reconcile the apparent contradictions observed in the literature, particularly on the superposition of heterophily and homophily in online political discussions.

Chapter 2

Data Acquisition and Preprocessing

In this chapter, we explore the foundation of our analysis: the Reddit political discussion dataset. We begin by providing a basic understanding of Reddit, the platform we get data from (Section 2.1). Next, we explore the Reddit Political Compass, a crucial element for capturing user ideology within the platform (Section 2.2). Section 2.3 describes the specific process used to collect data on Reddit relevant to our research goals.

Following this data collection process, Section 2.4 begins an exploratory data analysis of the collected dataset. Here, we will examine the distribution of posts over time (Section 2.4.1) to understand a relevant temporal frame. In addition, we will examine the distributions of user activity and popularity (Section 2.4.2) to identify potential factors influencing user interactions. This comprehensive exploration of the data provides valuable insights and guides the direction of our subsequent analysis.

2.1 A Reddit Overview

Reddit is a social news aggregator and discussion website consistently ranked among the top ten most visited websites worldwide[1]. This website groups its content into thousands of topical communities called *subreddits*, which are organized around different topics. Users can pseudonymously post *submissions* in such subreddits and *comment* on other submissions or comments in a tree structure that generates discussions. From now on, we will use the word *message* to indicate both a submission and a comment for generality.
Unlike other social media platforms, Reddit organizes its home page around subreddits rather than user-to-user relationships, allowing users to stay informed about current events, learn about new topics, and connect with people who share their interests.
People can post different types of content, such as text, links, images, and videos, and other members can reply and upvote or downvote them. Both posts and comments have a *score*[2], which is calculated by subtracting the number of downvotes from the number of upvotes. This number is slightly obfuscated ('fuzzed') to prevent manipulation by spammers[3].
Several features make Reddit a good platform for studying social interactions and user behavior. Here are some key aspects that contribute to its research suitability:

- **Rich Data Volume:** Reddit has a massive user base, generating vast quantities of data in the form of billions of submissions and comments, alongside millions of user profiles. This data volume allows for robust analysis, facilitating the uncovering nuanced patterns within user behavior and interactions.

[1] https://www.semrush.com/website/reddit.com/overview/

[2] https://www.reddit.com/wiki/faq/#wiki_how_is_a_submission.27s_score_determined.3F

[3] Vote fuzzing is a technique used to obscure upvote and downvote numbers to prevent spam and protect user privacy. For example, if five users upvoted the comment and three users downvoted it, the upvote/downvote numbers may say 23 upvotes and 21 downvotes, or 12 upvotes and 10 downvotes. The score is correct, but the vote totals are 'fuzzed'.

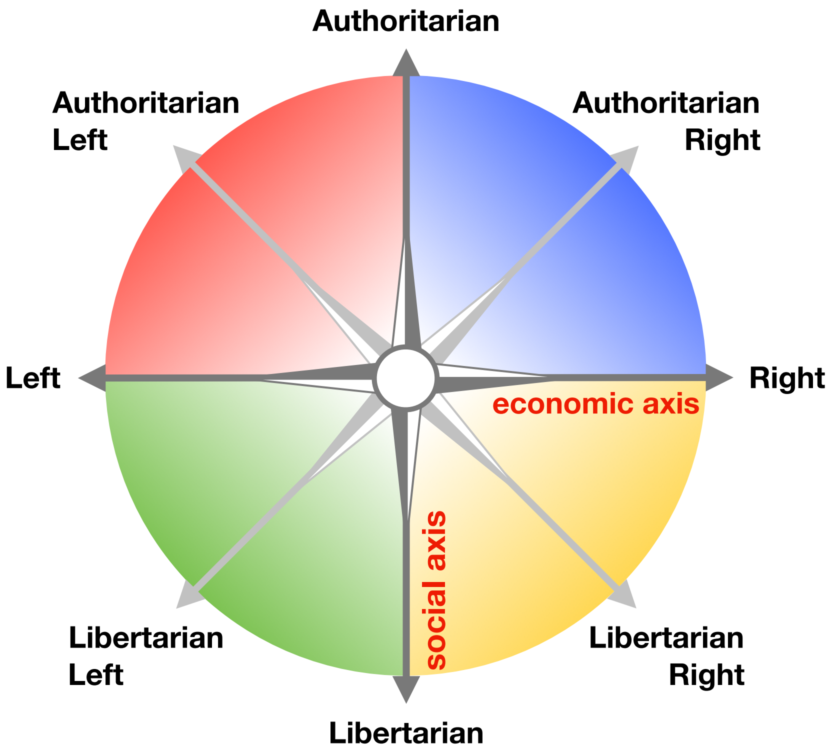


Figure 2.1: The political compass [42]: the horizontal axis delineates the economic ideologies, moving from left (equality-focused) to right (liberty-focused), and represents views on resource allocation. The vertical axis delineates the social ideologies, from libertarian at the bottom to authoritarian at the top, and represents views on personal freedom.

- **Network Structure:** The underlying structure of Reddit facilitates the construction of interaction networks. Submissions, comments, and user profiles naturally form connections that allow researchers to map and explore user interactions using network analysis techniques.
- **Anonymity and Free Expression:** While anonymity can pose some limitations, it also allows users to express themselves freely, potentially providing researchers with more authentic data compared to platforms with stricter identification requirements.

2.2 The Reddit Political Compass

Since we want to analyze the political discussions employing self-declared political ideologies, we pick two subreddits: `/r/PoliticalCompass` and `/r/PoliticalCompassMemes`.
`/r/PoliticalCompass`[4] is a community dedicated to posting and discussing test results, political self-tests, and political theory.
`/r/PoliticalCompassMemes`[5] is instead a humorous and satirical community where users share *memes* about politics and the political compass. Memes are cultural phenomena that spread rapidly online. They can take the form of humorous images with captions, or thought-provoking videos. The key is that they generate interest and are widely shared, becoming a common

[4] https://www.reddit.com/r/PoliticalCompass/
[5] https://www.reddit.com/r/PoliticalCompassMemes/

Table 2.1: List of all flairs expressing political ideologies in `/r/PC` and `/r/PCM`.

flair	Social Axis	Economic Axis
`AuthRight`	`Auth`	`Right`
`AuthCenter`	`Auth`	`Center`
`AuthLeft`	`Auth`	`Left`
`Right`	`Center`	`Right`
`Centrist`	`Center`	`Center`
`Left`	`Center`	`Left`
`LibRight`	`Lib`	`Right`
`LibCenter`	`Lib`	`Center`
`LibLeft`	`Lib`	`Left`

Table 2.2: JSON attributes of the Reddit dataset for submissions and comments, and their descriptions. `parent_id` and `body` are attributes only for comments JSON, while the others are common to both submissions and comments. There are valid attributes of JSON data for both `/r/PC` and `/r/PCM`.

JSON Attribute	Submission	Comment	Description
`id`	✓	✓	unique alphanumeric message identifier
`parent_id`	✗	✓	alphanumeric parent identifier of the current comment (identifier (`id`) of a message)
`author`	✓	✓	user pseudonym
`author_flair_text`	✓	✓	user political self-declaration
`body`	✗	✓	comment text
`subreddit`	✓	✓	message subreddit
`score`	✓	✓	message score
`created_utc`	✓	✓	message creation date

experience for many internet users. They can reflect current events, social trends, or inside jokes within a particular community.

On Reddit, users can customize their profiles by adding tags called 'flairs' to their usernames. Flairs vary from community to community. They serve two main purposes. First, they can provide context by offering additional information about a message posted by a user, clarifying the topic or purpose. Second, in subreddits specifically dedicated to political discussion, such as `/r/PC` and `/r/PCM`, flairs function as self-declared badges of political ideology. As shown in Figure 2.1, flairs are positioned on this political compass based on their economic and social attributes (Section 1.2.1). This allows us to analyze and categorize users based on their self-declared political identities through their chosen flairs. Since there are three positions for each axis, for a total of nine flairs, each flair represents a specific ideology within this two-dimensional space. For example, `Centrist` represents a neutral or balanced view on both economic and social issues, while `AuthLeft` reflects an ideology that combines authoritarian tendencies with a left-wing economic stance. Table 2.1 summarizes all the flairs expressing political ideologies present in `/r/PC` and `/r/PCM`.

2.3 Collecting Data on Reddit

We collect Reddit submissions and comments using Pushshift [10], a social media data collection, analysis, and archiving platform that has been collecting Reddit data and making it available to researchers since 2015. Each submission and comment is a JSON object, a file format and data exchange format that uses human-readable text to store and transfer data objects consisting of attribute-value pairs. We focus on a subset of these attributes, summarized in Table 2.2, useful for preliminary analysis and interaction network reconstruction.

As shown, `body` and `parent_id` are attributes only for comments; the latter is the fundamental element to reconstruct the interaction network. All the other attributes are common between submissions and comments. To restrict the analysis only to `/r/PC` and `/r/PCM`, we filter Reddit data to include only messages in those communities (through the `subreddit` attribute).
We filter out submissions and comments with null `id` and `parent_id`, since they are essential to reconstruct the interaction networks. We collect the `body` (of a comment) to analyze the language toxicity associated with that comment (Section 5.4), and the `created_utc` attribute to exploit the number of submissions and comments over time, helpful in selecting a specific period to focus on (Section 2.4.1).

2.4 Exploratory Data Analysis

To gain a comprehensive understanding of our dataset, we employ Exploratory Data Analysis, a necessary first step in any data analysis process. It is a set of techniques used to get a basic understanding of the data and uncover any interesting patterns, trends, or potential issues. These tools not only help us determine the size of the dataset (defining the scope of our analysis and its limitations) but also play a crucial role in identifying the starting period for user flair. Establishing a consistent timeframe for analyzing user interactions is essential for a meaningful analysis. In addition, we examine user activity and popularity distributions. These characteristics are important to consider because they can act as confounding factors, influencing user interactions independently of the relationships we are studying. Identifying such factors is critical to building robust statistical models that produce accurate and reliable results (Section 5.3).

2.4.1 Post Volume Over Time

Our analysis of the dataset, spanning from 2012 to 2022 for `/r/PC`, and from 2018 to 2022 for `/r/PCM`, reveals several key insights (Figure 2.2). Both subreddits showed little to no activity until 2018. More importantly, political ideologies have become a prominent feature only since November 2019, suggesting that focusing our analysis on the timeframe of 2020–2022 would be most productive.

The dataset comprises substantial data, capturing nearly all user-generated content from these subreddits. For `/r/PC`, it includes 79 368 submissions (96% of all submissions) and 952 550 comments (95% of all comments). Similarly, for `/r/PCM`, we have 383 169 submissions (96% of all submissions) and 22 653 346 comments (98% of all comments). This comprehensive coverage across both subreddits justifies our decision to focus on the 2020–2022 timeframe for our analysis.

For our analysis, we select real active users identified with a unique political ideology, thereby filtering out potential bots [55], deleted accounts, and those not affiliated with any political ideology or affiliated with multiple ideologies. Within the `/r/PC` dataset for 2020-2022, 22 503 users are aligned with a single ideology, whereas 2544 have multiple ideological affiliations. In `/r/PCM`, these numbers are 258 428 and 30 658 respectively. The breakdown of posts and comments by these user types is detailed in Table 2.3.

Figure 2.3 shows a comprehensive overview of user composition across both social and economic dimensions of political ideology, for both `/r/PC` and `/r/PCM`. On the social axis, libertarian ideology has a bigger representation of the centrist and mostly the authoritarian one.

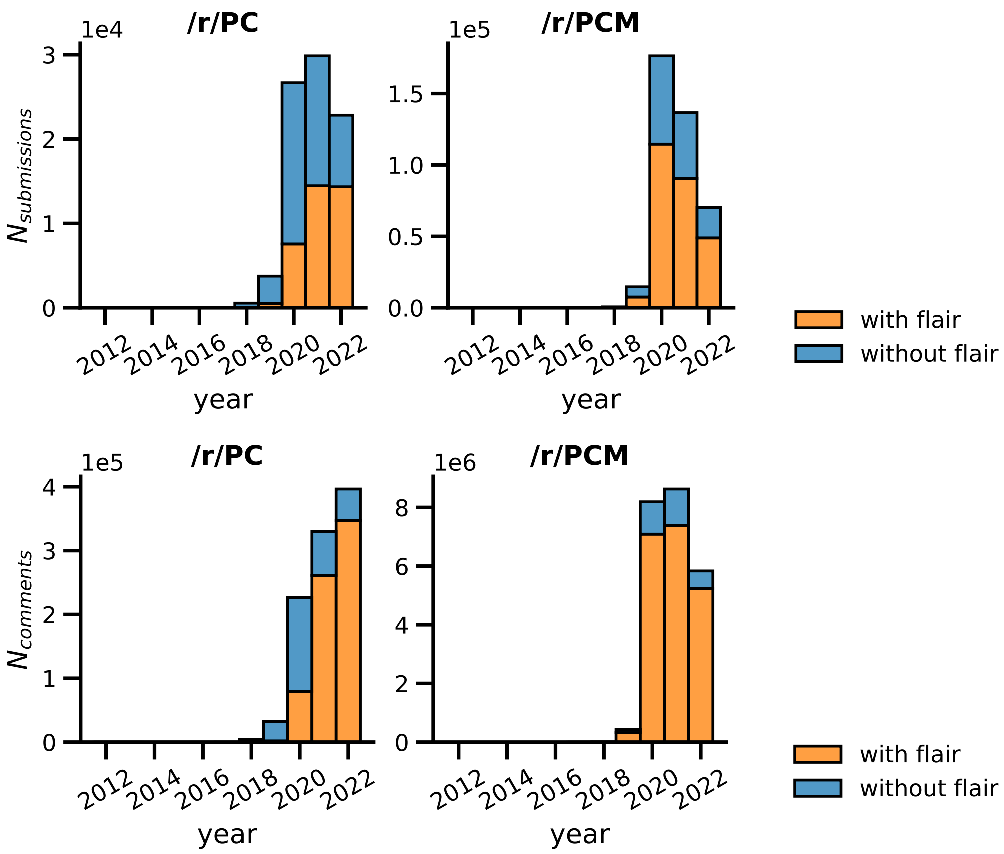


Figure 2.2: Number of submissions (top) and comments (bottom) per year in both subreddits.

Table 2.3: Fraction of posts and comments by user category in /r/PC and /r/PCM. The rows represent the following user types: deleted users, bots (as listed in [55]), users lacking a unique political declaration, and users with a unique political declaration. The last row indicates the fraction of analyzed data.

Subreddit	/r/PC		/r/PCM	
User Type	Posts	Comments	Posts	Comments
Deleted	34.72%	9.49%	30.25%	10.82%
Bots	0.25%	0.6%	0.73%	2.57%
Non-unique political flairs	28.41%	30.83%	21.98%	22.28%
Unique political flairs	**36.62%**	**59.08%**	**47.04%**	**64.33%**

On the economic axis, the left ideology is slightly more present than the right one. Both /r/PC and /r/PCM share these insights.

Figure 2.4 shows a breakdown of the number of submissions and comments on both dimensions of political ideology. Libertarians tend to be more active than authoritarians in submitting content and leaving comments. This trend is particularly noticeable on /r/PCM. On the other hand, left-leaning users are more active than right-leaning users on /r/PC. On /r/PCM, however, the trend reverses. Right-leaning users are much more active than left-leaning users. The data

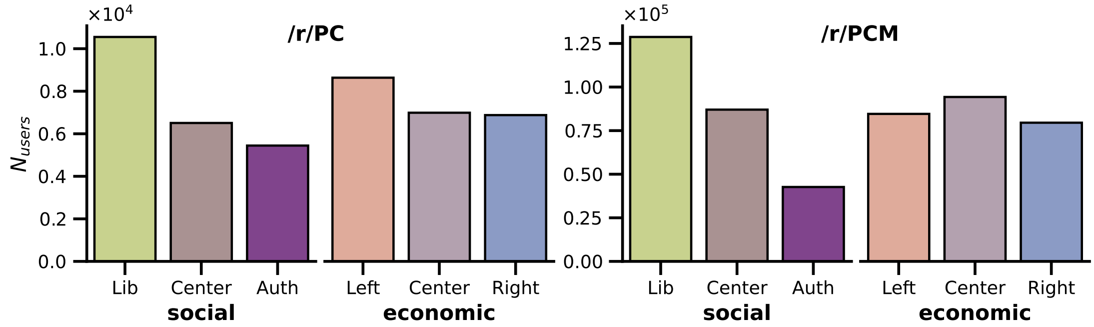


Figure 2.3: Composition of /r/PC and /r/PCM in terms of the number of users classified by political ideology.

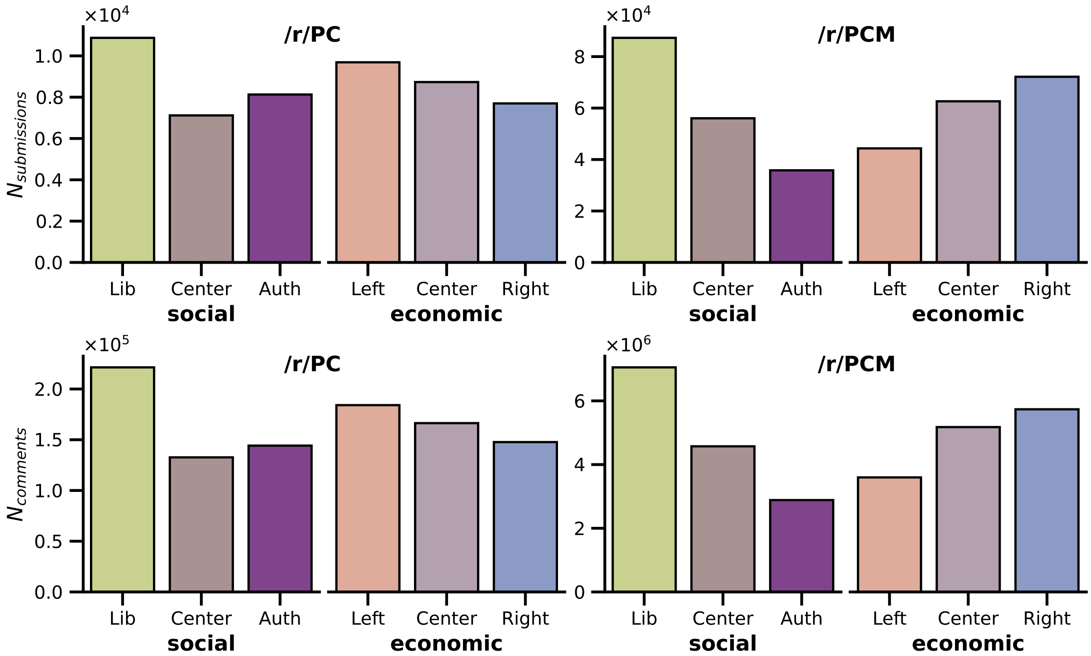


Figure 2.4: Composition of /r/PC and /r/PCM in terms of the number of submissions (top) and comments (bottom) by political ideology.

collected is representative of all classes, both in terms of users and activity, as shown in the previous plots.

2.4.2 Activity and Popularity Distributions

Exploring the dataset, we track the activity and the popularity of users. We define the **activity** a as the number of messages the user posted to the subreddit, and the **popularity** as the average score s calculated based on the score of each message the user posted to the subreddit.
Figure 2.5 shows the activity distribution $P(a)$ on the left: one can see that many users leave a small number of comments, while a few of them post a lot. Users on `/r/PCM` are more active than those on `/r/PC`.
Figure 2.6 shows how active users are on `/r/PC` and `/r/PCM` based on their social and economic ideologies. Authoritarian users comment more frequently, while right-wing users are slightly more active than left-wing users. This trend is more pronounced on `/r/PCM`. Interestingly, authoritarian and right-wing users make up a smaller portion of the total user base than libertarian and left-wing users, but they are more active.

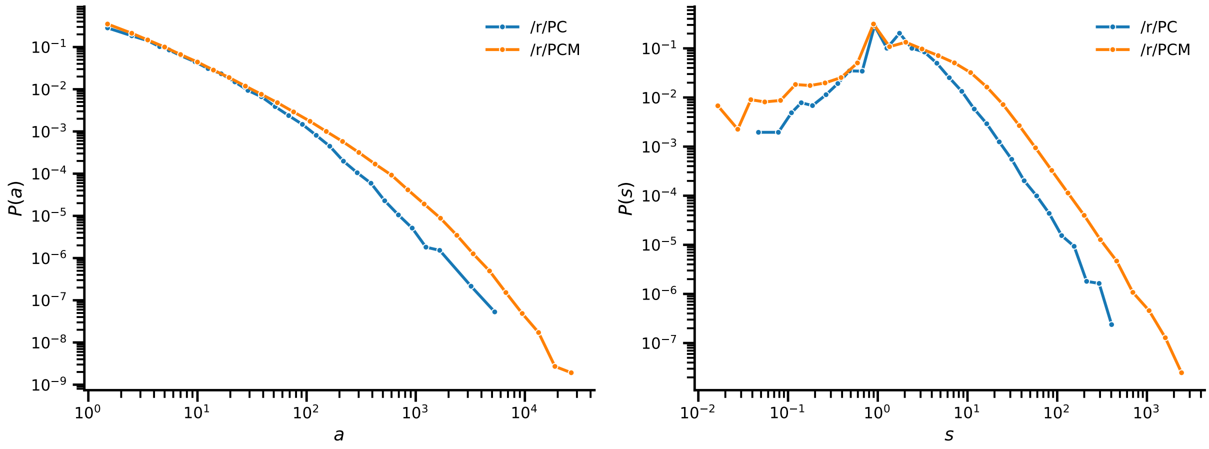
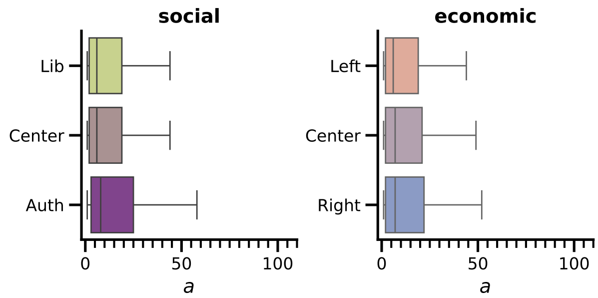
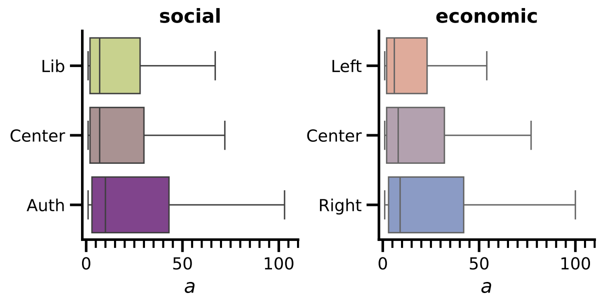


Figure 2.5: Probability distribution $P(a)$ of the activity a (left), and probability distribution $P(s)$ of the popularity s (right) of users plotted separately for `/r/PC` and `/r/PCM`.



(a) `/r/PC`.

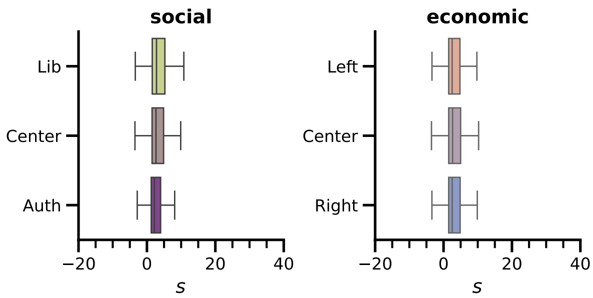


(b) `/r/PCM`.

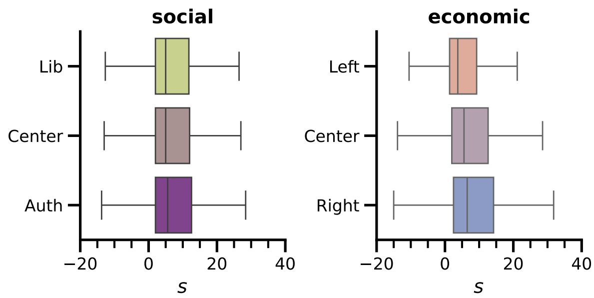
Figure 2.6: Activity (a number of comments per user) and popularity (average score s per user) distributions $P(a)$ for both subreddits.

Figure 2.5 displays the popularity distribution $P(s)$ on the right. Most users have a low level of popularity, meaning that the average score of their comments falls between 1 and 10. On `/r/PCM`, the popularity covers a wider range of values. A message's score can be negative, resulting in some users having a "negative" popularity due to the majority of their messages receiving negative scores.
The user popularity grouped by ideology in Figure 2.7 shows on the social axis libertarians are a little more popular than authoritarian ones. Their average score spans a little more wide interval. On the economic axis, there is little difference in popularity among economic ideologies. This holds for both `/r/PC` (Figure 2.7 a) and `/r/PCM` (Figure 2.7 b). However, popularity values are generally higher on `/r/PCM` compared to `/r/PC`, and the differences in popularity among ideologies on the two axes are more pronounced.

The popularity s is a useful element that can be utilized to recognize the potential influence of a user on Reddit, representing a confounding effect. The latter could influence the formation of interactions. This is the reason why we kept it and included it in the logistic regression model to analyze the feature importance in Section 5.3.1.



(a) /r/PC.



(b) /r/PCM.

Figure 2.7: Activity (a number of comments per user) and popularity (average score s per user) distributions $P(a)$ for both subreddits.

Chapter 3

Demographic Features

We get the user political ideology as a self-declaration from the `author_flair_text` JSON attribute, without having to infer it. To better characterize users in our dataset, we want to consider also demographic features—including age (old/young), gender (male/female), and affluence (poor/rich). The two employed subreddits don't give any chance to extract such data, so we need an unsupervised approach to infer them (Section 3.1). We employ the unsupervised representation learning approach by Waller and Anderson [65]. We then infer all of the previously listed demographic characteristics—age, gender, and affluence—and also get the partisanship (left/right) to see if this inferred dimension aligns with the economic ideologies obtained from user self-declarations. By incorporating these inferred features (Section 3.2), we can create a richer demographic profile for each user within our analysis.

3.1 Community Embedding and Demographic Scores

Building upon a community embedding technique, the unsupervised approach developed by Waller and Anderson [65] represents each Reddit community (subreddit) as a 150–dimensional vector. The latter captures the dominant topics and discussions within that community.
The method identifies a set of key communities, called "seeds," for each demographic feature. As illustrated in Figure 3.1, this is done through a three-step process:

1. **Finding Opposites**: The first step involves the search for a pair of communities that represent opposite ends of the spectrum for the target feature. For example, to analyze age, it might choose communities like `r/teenagers` and `r/RedditForGrownups`.

2. **Expanding the Set**: Next, the method searches for additional pairs of communities similar to the initial one. This helps capture the range and variations within the feature. Continuing the age example, this might include communities like `r/highschools` and `r/AskMenOver30`.

3. **Finding the Balance**: Finally, the method combines information from all the seed communities. This creates a single score for each community on the target feature. Communities

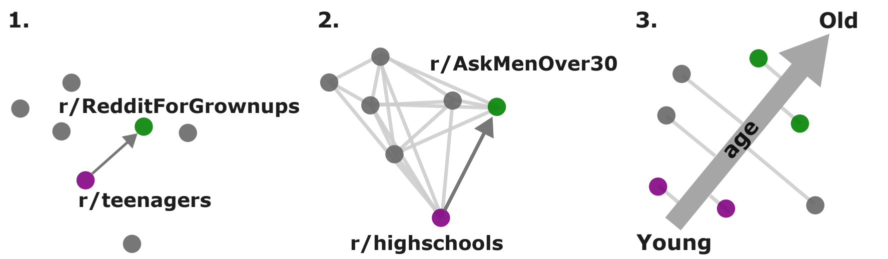


Figure 3.1: An illustration of the methodology to generate age feature by Waller and Anderson [65].

Table 3.1: Fraction of users and interactions in /r/PC and /r/PCM due to activity within the set of subreddits used for scoring.

	/r/PC	/r/PCM
users	81.13%	97.26%
interactions	95.15%	96.36%

very similar to one "polar" seed will have a high score, while those in the middle will receive a score closer to zero.

Communities that are more aligned with one end of the spectrum (e.g., topics or discussions typically associated probably with young users) will receive a higher score on that end. The researchers applied this approach to analyze the 10 006 most active Reddit communities (defined as $\mathcal{S}$, the set of subreddits used for scoring) based on user activity between 2005 and 2018.

3.2 Inferring Socio-Demographics

To leverage effectively the scores for the demographic features of Reddit communities introduced in Section 3.1, we need to analyze user activity within the same set of subreddits $\mathcal{S}$. In other words, we need to ensure that most user activity occurs in the communities we have scores for. Fortunately, our analysis aligns well with this requirement. The user activity in terms of comments number in subreddits $s \in \mathcal{S}$ is 56.94% and 68.02% respectively for /r/PC and /r/PCM. The vast majority of users actively commented in these subreddits. A small number of users did not comment on any of these subreddits. Due to the lack of data, we cannot infer their demographic characteristics and exclude them from further analysis. However, this exclusion has a low impact on the overall data, since their contribution is little, as shown in Table 3.1.

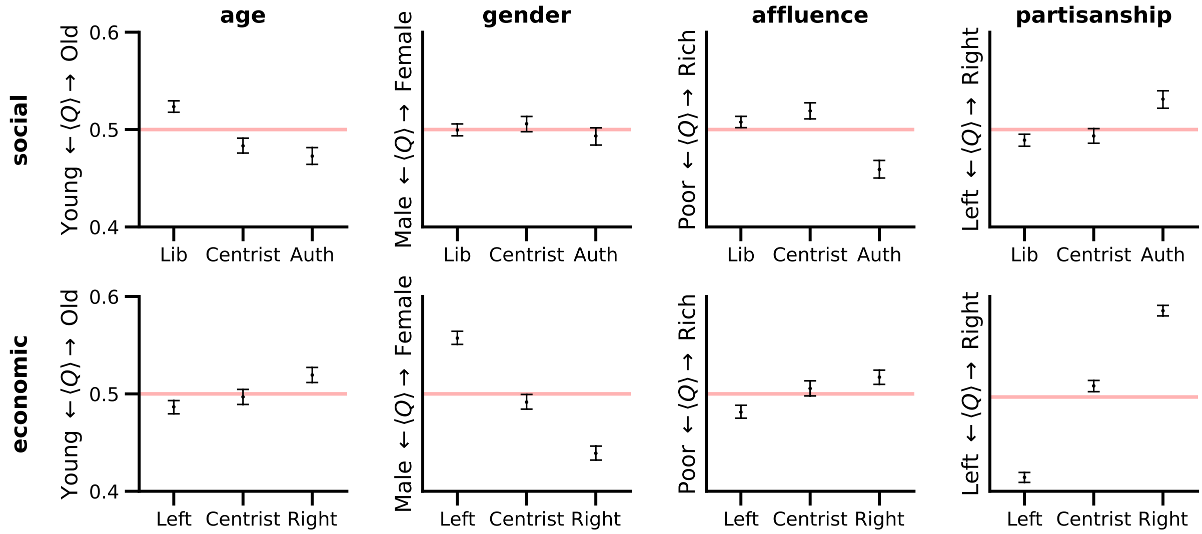
To determine the z-score for each user u based on a characteristic c from the set {age, gender, affluence, partisanship}, $z_u^{(c)}$, we compute a weighted mean of the z-scores $z_s^{(c)}$ of all subreddits $s \in \mathcal{S}$. The weight is determined by the number of comments $N_{u,s}$ that user u posted in subreddit s. The weighted average is thus:

$$z_u^{(c)} = \frac{\sum_{s \in \mathcal{S}} N_{u,s} z_s^{(c)}}{\sum_{s \in \mathcal{S}} N_{u,s}}. \quad (3.1)$$

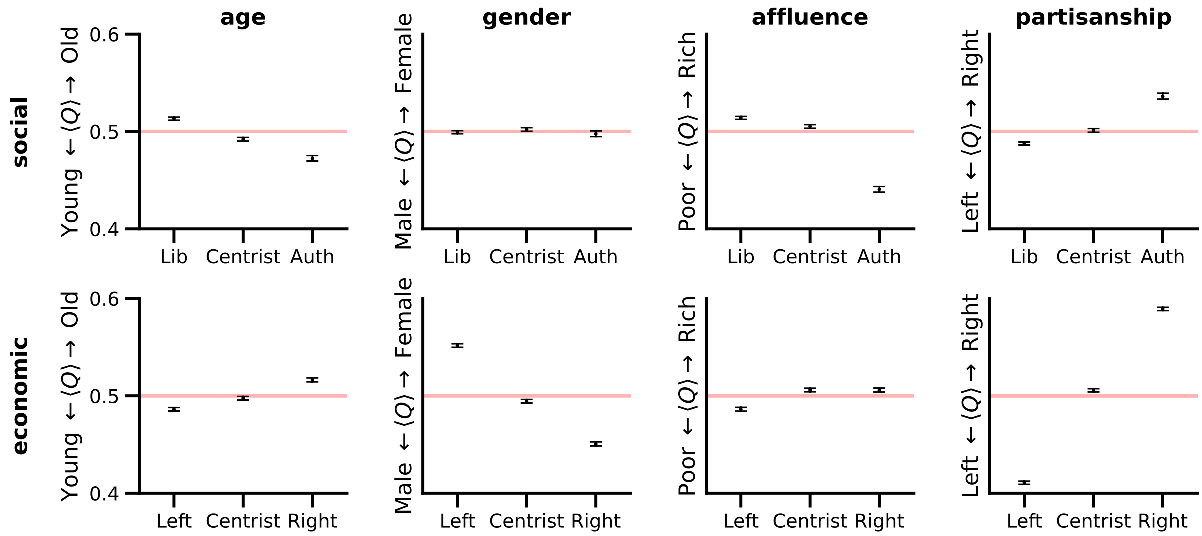
Subsequently, we normalize the z-scores of users by using quantiles for each characteristic. These normalized scores will henceforth be referred to as *quantile scores* (Q). Quantile scores in the top 25% are classified as "high", those in the bottom 25% as "low", and quantile scores in between belong to the reference class. For instance, within the age characteristic, a high score would correspond to an "old" user, while a low score represents a "young" one.

We stress that a score indicating an old-leaning user does not imply that such a user is necessarily old, as Reddit is known to be participated by more young than old users [7]. Rather, it indicates a user who frequents subreddits more likely to be participated by older users. A similar argument holds for the other categories: demographic attributes are always to be considered relative to the Reddit user base, and never on an absolute basis.

Figure 3.2 illustrates the distribution of demographic characteristics among different political ideologies for /r/PC (a) and /r/PCM (a). In each plot, the y-axis represents the average quantile score of a demographic characteristic for users with the corresponding ideology on the x-axis. Notably, the distribution of inferred ideologies (left/right) closely mirrors the declared ideologies from the Political Compass along the economic axis, lending validation to our analytical approach. This result holds for both /r/PC and /r/PCM.



(a) `/r/PC`.



(b) `/r/PCM`.

Figure 3.2: For each axis (social and economic) displayed in the rows, and for each user characteristic (age, gender, affluence, partisanship) displayed in the columns, the plot shows the average quantile score and its 95% confidence intervals for `/r/PC` (a) and `/r/PCM` (b). Libertarians tend to be older, while right-leaning users are predominantly male. Left-leaning users are more likely to be female. Additionally, a correlation between partisanship and the left–right economic axis emerges, which further validates our data collection methodology. Both `/r/PC` and `/r/PCM` have the same results.

Figure 3.2 reveals noticeable variations in the characteristics of Reddit users based on their diverse socio-demographic attributes and political ideologies. Libertarians are typically older, wealthier, and more left-leaning compared to authoritarians.

On the contrary, left-wing users tend to be younger and have a higher female representation but are generally less rich compared to right-wing users. Such results hold also for `/r/PCM`, as shown in Figure 3.2b.

Chapter 4

Modeling Multidimensional Ideologies

While the previous chapters focused on data acquisition and user characteristics, this chapter delves into the core of our investigation: modeling multidimensional ideologies and user interactions on Reddit. We begin by reconstructing the interaction network in Section 4.1, by mapping the connections between users, revealing how information and ideas spread across the network. Then, in Section 4.2, we explore the probabilities of interaction, by considering both joint and conditional probabilities. This analysis allows us to quantify the likelihood of user interactions based on their ideological positions. Finally, Section 4.3 establishes a null model for the social network, providing a baseline for comparison and helping us assess the significance of the observed interaction patterns. This chapter explores the relationship between user ideologies and interactions in political discussion. The goal is to understand how a user's ideological position affects their interactions with others.

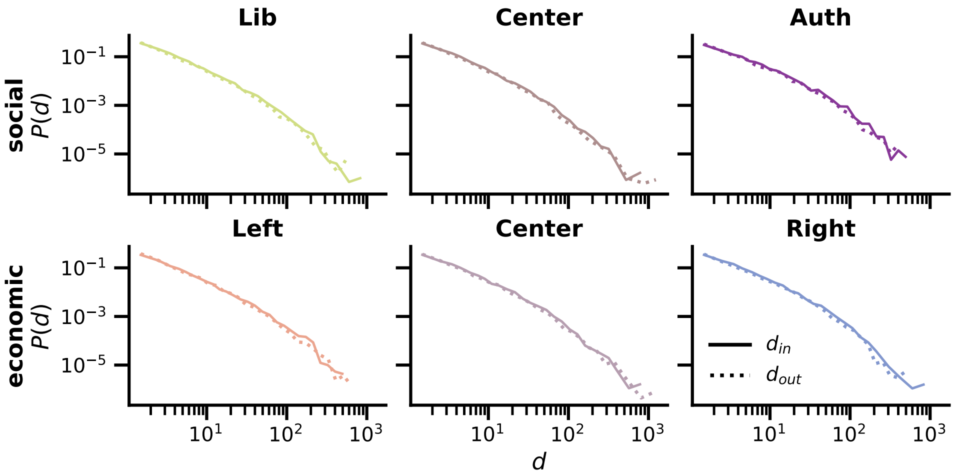
4.1 Reconstructing the Interaction Network

Social network analysis relies heavily on graphs to visualize and analyze the connections between individuals. These graphs consist of nodes, which represent individual users, and edges, which represent the interactions or relationships between them. Since the key question is to seek for tendencies of users to connect with similar (homophily) or dissimilar (heterophily) others, the direction analysis of messages can shed light on this.
A directed graph offers a significant advantage over an undirected graph. Unlike an undirected graph where an edge simply signifies a connection, a directed graph reveals the direction of communication, specifying the source and the target of the interaction. We can see if users mainly send messages to those with similar views or engage in cross-ideological communication. This directionality allows us to capture the flow of interaction and identify patterns of influence within the network. It establishes the fundamental element needed to explore homophily and heterophily in online discussions. Additionally, we can quantify the extent of these effects.
Furthermore, graphs can incorporate weight on their edges. This weight goes beyond simply acknowledging a connection and instead assigns a value that reflects the strength or intensity of the interaction, i.e. the number of messages exchanged. By incorporating weight, we gain a richer understanding of the social network, identifying not just "who connects with whom", but also the "strength" of those connections.

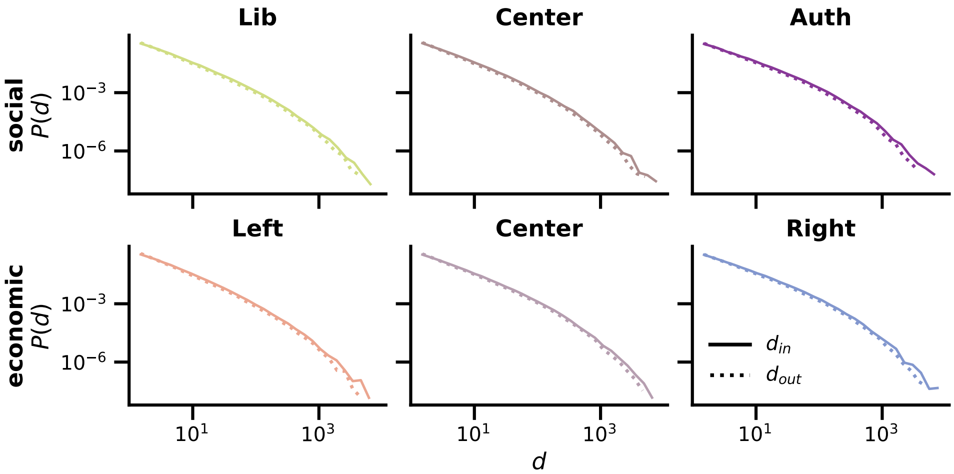
We represent the interaction network as a directed weighted graph $G = (V, E, w)$, where users are nodes V and the edges E correspond to interactions between them. Here, we focus on interactions where a user replies to another, and therefore, we exclude self-edges (i.e., a user replying to their comment).
An edge $(u, v) \in E$ indicates that user u (source) replied to user v (target) in a thread on Reddit. Each edge carries a weight w_{uv} which reflects the number of interactions between u and v. Table 4.1 summarizes the number of nodes $|V|$, the number of edges $|E|$, the average degree $\langle d \rangle$ and the total number of interactions W, for both `/r/PC` and `/r/PCM`. Figure 4.1 shows instead

Table 4.1: Network statistics: number of users $|V|$, edges $|E|$, average degree $\langle d \rangle$, and total number of interactions W.

Subreddit	$\|V\|$	$\|E\|$	$\langle d \rangle$	W
`/r/PC`	18 135	173 672	9.58	261 078
`/r/PCM`	215 111	6 197 901	28.81	8 065 395



(a) `/r/PC`.

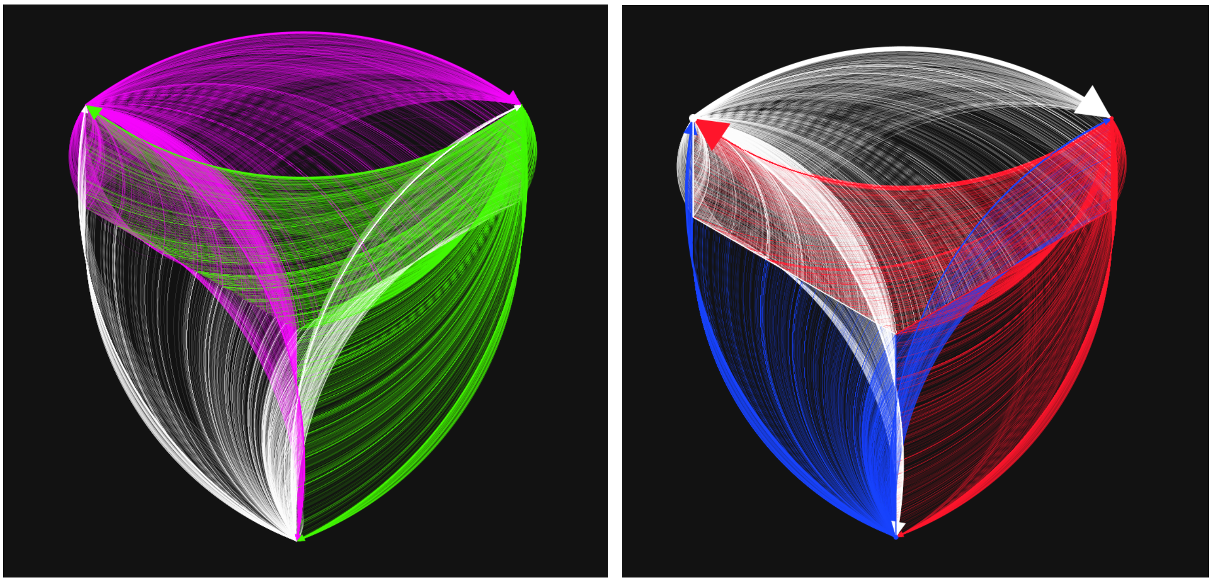


(b) `/r/PCM`.

Figure 4.1: The in-degree and out-degree distributions $P(d)$ (with $d \in \{d_{in}, d_{out}\}$) of nodes for each ideological axis (rows) and ideology (columns), for both `/r/PC` (a), and `/r/PCM` (b).

the in- and out-degree distributions of the nodes of the networks, grouped by ideologies of the social and economic axes.

We establish the directionality of the graph by using two key attributes within each comment (as defined in Table 2.2): `id` and `parent_id`. When user u replies to user v, essentially creating a directed edge denoted as $u \to v = (u, v) \in E$, the comment posted by user u has a `parent_id` attribute that matches the `id` attribute of the message posted by user v. This connection between the `parent_id` and the original `id` allows us to identify the direction of the reply, hence the



(a) (b)

Figure 4.2: Hive plots showing interactions between ideologies in `/r/PC` for social (a: 'violet' for authoritarian, 'white' for centrist, 'green' for libertarian) and economic (b: 'blue' for right, 'white' for centrist, 'red' for left) axes. The size of each node represents the out-degree of nodes within each ideology group. The edges connecting the circles depict the interactions between nodes of different ideologies.

Table 4.2: Marginal probabilities to find a node labeled as $X_s \in \{B, C, A\}$ and $X_e \in \{L, C, R\}$ respectively for social and economic axis for both `/r/PC` and `/r/PCM`.

Ideological Axis	social			economic		
Subreddit	$P(B)$	$P(C)$	$P(A)$	$P(L)$	$P(C)$	$P(R)$
`/r/PC`	0.48	0.30	0.23	0.38	0.31	0.31
`/r/PCM`	0.51	0.32	0.17	0.35	0.34	0.31

direction of the edge in the graph.
Figure 4.2 shows the hive plots for `/r/PC`, for both ideological axes, while Figure 4.3 shows a graphical representation of the network reconstruction process.

4.2 Interaction Probabilities

For a user with ideology X_s from {libertarian, center, authoritarian} ($\{B, C, A\}$ for simplicity) on the social axis (s) and X_e from {left, center, right} ($\{L, C, R\}$) on the economic axis (e), the probability of observing them is given by $P(u = X) = N_X/|V|$, where $|V|$ is the total number of users, and N_X stands for those users identified with the ideology X. In particular, the Table 4.2 shows the probabilities to find a node labeled as $X_s \in \{B, C, A\}$ and $X_e \in \{L, C, R\}$ for both `/r/PC` and `/r/PCM`.

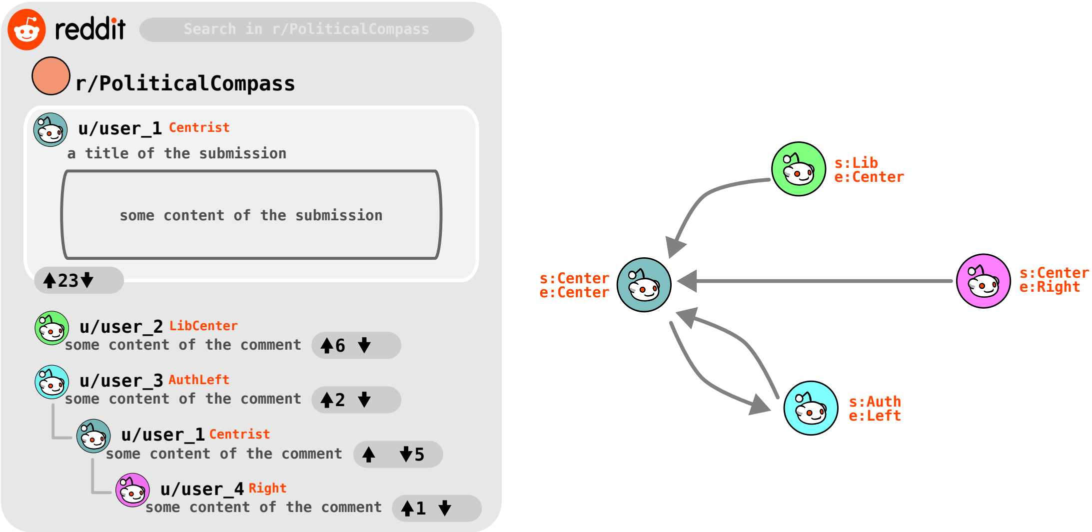


Figure 4.3: An illustration of the reconstructing process of the network from user interactions. On the left side, we see an example of a discussion in /r/PC between four users. On the right side, the reconstructed network is shown. This network depicts the connections between users and their ideologies on both social and economic axes.

4.2.1 Joint Probabilities

When considering a specific ideological axis, either social or economic, the joint interaction probability between a user of ideology X and another of ideology Y is $P_{X\to Y} = \frac{W_{X\to Y}}{W}$, where W is the total number of interactions in the network, and

$$W_{X\to Y} = \sum_{u,v\in V: u=X \wedge v=Y} w_{u,v}$$

is the total weight of directed edges from X to Y. Empirically, in /r/PC we find:

$$P_{X_s\to Y_s} \simeq \begin{array}{c c} & \begin{array}{ccc} B & C & A \end{array} \\ \begin{array}{c} B \\ C \\ A \end{array} & \begin{pmatrix} 0.21 & 0.11 & 0.13 \\ 0.11 & 0.08 & 0.08 \\ 0.12 & 0.07 & 0.10 \end{pmatrix} \end{array}, \quad P_{X_e\to Y_e} \simeq \begin{array}{c c} & \begin{array}{ccc} L & C & R \end{array} \\ \begin{array}{c} L \\ C \\ R \end{array} & \begin{pmatrix} 0.13 & 0.12 & 0.12 \\ 0.12 & 0.12 & 0.09 \\ 0.12 & 0.09 & 0.08 \end{pmatrix} \end{array}.$$

In /r/PCM we get:

$$P_{X_s\to Y_s} \simeq \begin{array}{c c} & \begin{array}{ccc} B & C & A \end{array} \\ \begin{array}{c} B \\ C \\ A \end{array} & \begin{pmatrix} 0.24 & 0.15 & 0.10 \\ 0.15 & 0.10 & 0.06 \\ 0.09 & 0.06 & 0.05 \end{pmatrix} \end{array}, \quad P_{X_e\to Y_e} \simeq \begin{array}{c c} & \begin{array}{ccc} L & C & R \end{array} \\ \begin{array}{c} L \\ C \\ R \end{array} & \begin{pmatrix} 0.06 & 0.09 & 0.12 \\ 0.09 & 0.12 & 0.13 \\ 0.12 & 0.13 & 0.14 \end{pmatrix} \end{array}.$$

Independent of the ideological axis, the diagonal elements of the previous matrices correspond to the interactions within ideologies, off-diagonal to those across ideologies. The sum by rows corresponds to the probability that a node with ideology X initiates an interaction $P_{X\to} = W_{X\to}/W$. The sum by columns corresponds instead to the probability that a node with ideology X receives an interaction $P_{\to X} = W_{\to X}/W$. Observing the matrices, one can notice that cross-interactions look almost symmetric between the different ideological groups, both for social and economic

axes.
However, joint probabilities do not take into account the difference in size between the different groups. This can be observed, for example, by noticing that the probability an authoritarian user can start an interaction is given by $P_{A\to} \simeq 0.29$, larger than the marginal probability of having an authoritarian user $P(A) \simeq 0.23$, with $P(A) < P_{A\to}$. The conditional probabilities can provide a more detailed examination of these characteristics.

4.2.2 Conditional Probabilities

Let $W_{\to Y} = \sum_X W_{X\to Y}$ be the number of interactions received by Y, we consider the conditional probability of getting an interaction $X \to Y$ given a source with ideology X as:

$$P_{X\to Y|X} = \frac{P_{X\to Y}}{P_{X\to}} = \frac{W_{X\to Y}}{W_{X\to}}. \tag{4.1}$$

Empirically, for `/r/PC`, we have:

$$P_{X_s\to Y_s|X_s} \simeq \begin{array}{c} \\ B \\ C \\ A \end{array} \begin{array}{c} \begin{array}{ccc} B & C & A \end{array} \\ \begin{pmatrix} 0.47 & 0.25 & 0.28 \\ 0.42 & 0.28 & 0.29 \\ 0.41 & 0.25 & 0.34 \end{pmatrix} \end{array}, \tag{4.2a}$$

$$P_{X_e\to Y_e|X_e} \simeq \begin{array}{c} \\ L \\ C \\ R \end{array} \begin{array}{c} \begin{array}{ccc} L & C & R \end{array} \\ \begin{pmatrix} 0.36 & 0.32 & 0.32 \\ 0.37 & 0.35 & 0.27 \\ 0.42 & 0.29 & 0.29 \end{pmatrix} \end{array}. \tag{4.2b}$$

For `/r/PCM` instead, we have empirically:

$$P_{X_s\to Y_s|X_s} \simeq \begin{array}{c} \\ B \\ C \\ A \end{array} \begin{array}{c} \begin{array}{ccc} B & C & A \end{array} \\ \begin{pmatrix} 0.50 & 0.30 & 0.20 \\ 0.48 & 0.32 & 0.21 \\ 0.46 & 0.29 & 0.25 \end{pmatrix} \end{array}, \tag{4.3a}$$

$$P_{X_e\to Y_e|X_e} \simeq \begin{array}{c} \\ L \\ C \\ R \end{array} \begin{array}{c} \begin{array}{ccc} L & C & R \end{array} \\ \begin{pmatrix} 0.22 & 0.33 & 0.45 \\ 0.26 & 0.35 & 0.38 \\ 0.31 & 0.33 & 0.36 \end{pmatrix} \end{array}. \tag{4.3b}$$

If people interact with each other irrespective of their ideology, we would expect everyone to have an equal chance of interacting with each group, depending only on the group size. However, we observe that this is not the case. Instead, Equations (4.2a) and (4.2b) show important deviations from this expectation, in addition to differences between the two ideological axes.

Interaction Patterns

On the **social axis**, the interactions tend to be more **homophilic** than heterophilic (the on-diagonal entry has a larger weight in each column). This fact is especially evident between libertarians and authoritarians: for `/r/PC` (Equation (4.2a)) authoritarians engage more with authoritarians (34%) compared to how much libertarians do (28%), and a similar pattern is observed in the opposite direction for libertarians (47% vs. 41%). Centrists also show reduced

interactions with both groups and favor instead their own group.
For `/r/PCM` (Equation (4.3a) the same qualitative results hold: authoritarians engage more with authoritarians (25%) compared to how much libertarians do (20%), and a similar pattern is observed in the opposite direction for libertarians (50% vs. 46%). Centrists also show reduced interactions with both groups and favor instead their own group.

Surprisingly, on the **economic axis** the trend inverts. There is a noticeable **heterophily** between left and right users (the off-diagonal entries have a larger weight in the respective columns). In `/r/PC` (Equation (4.2b)) left users receive more interaction from right ones (42%) than from within their group (36%), and the opposite holds true for right users (29% from the right vs. 32% from the left).
For `/r/PCM` (Equation (4.3b)) the same qualitative results hold: left users receive more interaction from right ones (31%) than from within their group (22%), and the opposite holds true for right users (36% from the right vs. 45% from the left).

After examining how conditional probabilities expose user interaction patterns based on ideological positions, a significant question arises: *"Are these observed patterns genuinely influenced by users' ideologies, or could they be a result of chance?"*
To address this question, Section 4.3 introduces the concept of a null model for the social network. This null model represents a random network that fulfills certain properties, and where user interactions are not influenced by ideology. By comparing the interaction patterns in our empirical network, which are derived from conditional probabilities, with those of a random network, we can assess the statistical significance of the observed patterns. This comparison will help us determine whether the observed interactions are likely due to the influence of user ideologies or simply random chance within the online discussion community.

4.3 Null Model for Social Network

To assess the statistical significance of the patterns observed in the interaction network (Equations (4.2a) and (4.2b) and Equations (4.3a) and (4.3b) in Section 4.2.2), we need to compare it with a null model. A random network offers simplicity but often fails to capture the non-random nature of real-world networks where users tend to connect with others who share similar characteristics. Therefore, we require a null model that disregards user ideology while preserving user activity.

We employ a configuration model [48], a directed, weighted random network (RN) that preserves the in-degree and out-degree sequences of the original network. This model rewires connections among nodes while ensuring that:

i user political ideology does not influence interaction patterns;

ii user activity (number of posted comments) and attractiveness (the number of received comments) are properly maintained.

To achieve these goals, we define a balanced sequence of pairs, i.e. we sample a number of negative examples equal to the positive ones. We first include all existing edges E from the empirical network. Then to get the negative examples we select a node u with a probability proportional to its activity, and a node v with a probability proportional to its attractiveness. If the obtained source–target pair $(u, v) \in E$ we discard it, otherwise we take it as a negative example. Algorithm 1 shows the approach just described.
Such a null model reflects the probability of considering a node pair as the product of two independent probabilities: the probability of the node u initiating an interaction, and the probability of the node v receiving an interaction.
In this RN model, the conditional probability of getting an interaction $X \to Y$, given the source's

Algorithm 1: Sampling negative edges.

Input: Edge list E
Output: Non-edge list E'
$E' \leftarrow \emptyset$;
while $|E'| < |E|$ **do**
    $s \leftarrow \min(2^{17}, |E| - |E'|)$;
    `//` 2^{17} `to prevent overflow`
    for $i \leftarrow 1$ **to** s **do**
        randomly select source u_i among E with the uniform probability;
        randomly select target v_i among E with the uniform probability;
        if $(u_i, v_i) \notin E$ **then**
            $E' \leftarrow E' \cup (u_i, v_i)$;
return E'

ideology X in the null model is independent of the class X of the source. Instead, it depends on the in-degree of the target's class Y and it is given as:

$$P^{RN}_{X\to Y|X} = \frac{P^{RN}_{X\to Y}}{P_{X\to}} = \frac{W_{\to Y}}{W}. \tag{4.4}$$

For the `/r/PC` dataset, the conditional probabilities for the social and economic axes are:

$$P^{RN}_{X_s\to Y_s|X_s} \simeq \begin{array}{c c} & \begin{array}{ccc} B & C & A \end{array} \\ \begin{array}{c} B \\ C \\ A \end{array} & \begin{pmatrix} 0.44 & 0.26 & 0.30 \\ 0.44 & 0.26 & 0.30 \\ 0.44 & 0.26 & 0.30 \end{pmatrix} \end{array}, \tag{4.5a}$$

$$P^{RN}_{X_e\to Y_e|X_e} \simeq \begin{array}{c c} & \begin{array}{ccc} L & C & R \end{array} \\ \begin{array}{c} L \\ C \\ R \end{array} & \begin{pmatrix} 0.38 & 0.32 & 0.30 \\ 0.38 & 0.32 & 0.30 \\ 0.38 & 0.32 & 0.30 \end{pmatrix} \end{array}. \tag{4.5b}$$

For the `/r/PCM` dataset, the conditional probabilities for the social and economic axes are instead:

$$P^{RN}_{X_s\to Y_s|X_s} \simeq \begin{array}{c c} & \begin{array}{ccc} B & C & A \end{array} \\ \begin{array}{c} B \\ C \\ A \end{array} & \begin{pmatrix} 0.48 & 0.30 & 0.21 \\ 0.48 & 0.30 & 0.21 \\ 0.48 & 0.30 & 0.21 \end{pmatrix} \end{array}, \tag{4.6a}$$

$$P^{RN}_{X_e\to Y_e|X_e} \simeq \begin{array}{c c} & \begin{array}{ccc} L & C & R \end{array} \\ \begin{array}{c} L \\ C \\ R \end{array} & \begin{pmatrix} 0.27 & 0.34 & 0.39 \\ 0.27 & 0.34 & 0.39 \\ 0.27 & 0.34 & 0.39 \end{pmatrix} \end{array}. \tag{4.6b}$$

Chapter 5

Results

This chapter presents the main findings and analyses of our study of online interactions. We begin by analyzing the influence of ideological factors on these interactions (Section 5.1). Here we examine both homophily (preference for interacting with similar ideologies) within social ideologies and heterophily (interaction across different ideologies), with a particular focus on economic attitudes (Sections 5.1.1 and 5.1.2).

Next, Section 5.2 examines the effect of demographics on the interactions. We use a logistic regression model (Section 5.2.1) to understand the relationship between users' demographics and their online interactions. We then examine homophily within demographic groups (Section 5.2.2).

Building on these individual analyses, Section 5.3 examines the combined effects of ideology, demographics, and potential confounders. We use a logistic regression model to assess the relative importance of these features (Section 5.3.1). This section goes on to examine specific aspects such as economic heterophily in addition to social homophily and demographic homophily.

Finally, Section 5.4 focuses on the analysis of language toxicity within these social interactions. We first establish a null model to serve as a baseline for comparison (Section 5.4.1), and then examine the actual levels of toxicity observed within the interactions (Section 5.4.2).

5.1 Analyzing the Ideological Effects on Interactions

We aim to validate the interaction patterns suggested by the conditional probability matrices calculated in Section 4.2 (Equations (4.2a) and (4.2b) for `/r/PC` and Equations (4.3a) and (4.3b) for `/r/PCM`). We achieve this by comparing these conditional probabilities with those arising from the null model (RN) introduced in Section 4.3.

To quantify the strength of observed interaction patterns based on user ideology, we employ the concept of *odds ratio* (OR). An OR measures how likely an event A occurs compared to its non-occurrence. If $P(A)$ represents the probability of event A happening, then the *odds* of event A are defined by:

$$OR(A) = \frac{P(A)}{1 - P(A)}.$$

While related, probabilities and odds differ in their ranges. Probability values ($P(A)$) lie between 0 and 1 ($0 \leq P(A) \leq 1$), while ORs can take on any value. Here are three key interpretations of OR values:

- $OR < 1$: Event A is less likely to occur than not occur.
- $OR = 1$: Event A has the same probability of occurring as not occurring (random chance).
- $OR > 1$: Event A is more likely to occur than not occur.

In our context, the event of interest is a directed interaction $X \rightarrow Y$, from ideology X to ideology Y, where X represents the ideology of the source node. The probability of observing this event is denoted by $P_{X \rightarrow Y|X}$. Our goal is to compare whether user ideology or randomness drives this interaction. Therefore, we compare the conditional probability of interaction ($P_{X \rightarrow Y|X}$) in the actual network with the corresponding probability in the null model ($P^{RN}_{X \rightarrow Y|X}$). The odds ratio for this comparison is defined as:

$$OR(X \rightarrow Y|X) = \frac{P_{X \rightarrow Y|X}}{P^{RN}_{X \rightarrow Y|X}}. \tag{5.1}$$

The OR value allows us to interpret user interaction patterns based on ideology:

- $OR(X \rightarrow Y|X) < 1$: Users with ideology X interact less with users of ideology Y compared to the null model, indicating heterophily.
- $OR(X \rightarrow Y|X) = 1$: No difference between the actual network and the null model suggests that randomness drives the interaction.
- $OR(X \rightarrow Y|X) > 1$: Users with ideology X interact more with users of ideology Y compared to the null model, indicating homophily.

5.1.1 Homophily within Social Ideologies

Figure 5.1 shows the odds ratios between empirical and random conditional probabilities of interaction for both `/r/PC` and `/r/PCM`, according to Equation (5.1).
Figure 5.1a illustrates that political ideologies on the social axis (libertarian–authoritarian) for `/r/PC` exhibit significant homophily. Interactions within the same ideology (on-diagonal) are up to 12% more likely than those predicted by the null model. The interaction probabilities both in receiving and sending a comment from/to a differing ideology are approximately symmetric.
This pattern holds for `/r/PCM` too, as shown in Figure 5.1b, where interactions between authoritarian users are up to 18% more likely than expected.

5.1.2 Heterophily across Economic Ideologies

Conversely, the right side of Figure 5.1a shows a pronounced heterophily across ideologies on the economic axes. Left- and right-leaning users are approximately 10% more likely to interact with each other than what the null model predicts. This phenomenon of increased interaction between left and right is even more prominent in the `/r/PCM` subreddit, where is observed 15% more likely than predicted, see Figure 5.1b.

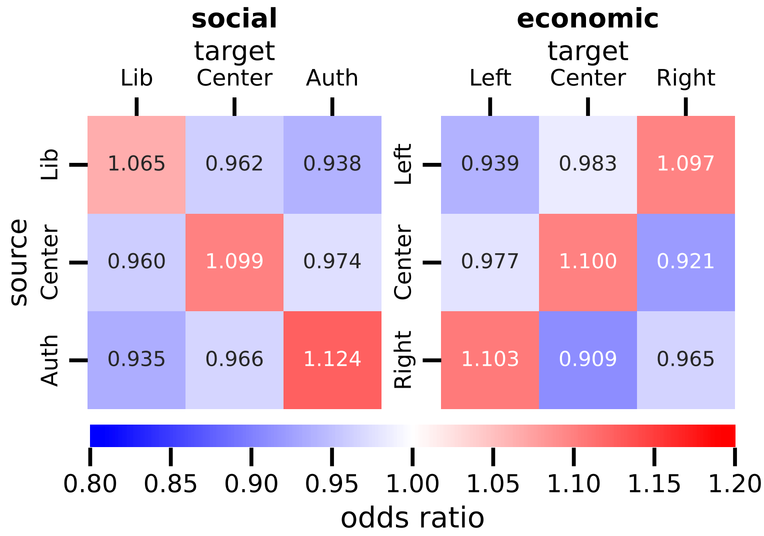
Furthermore, within-group interactions for left and right are considerably lower than expected, while the economic center group shows more within-group interactions.

5.2 Analyzing the Demographics Effects on Interactions

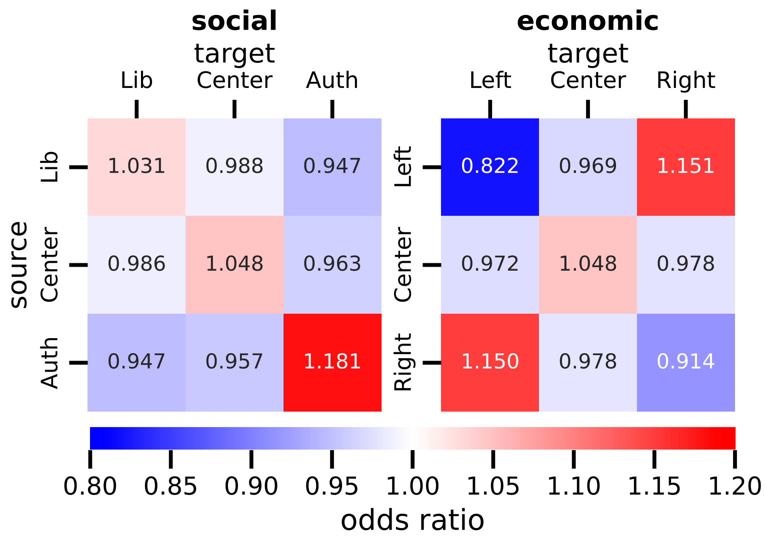
Next, we test how social interactions are associated with demographic factors. To this aim, we employ the logistic regression model by Monti et al. [49] which outputs the probability of interaction between demographic groups and validates the statistical significance of the results.

5.2.1 Logistic Regression Model

A logistic regression model can be used to understand the relationship between various user features and a particular outcome of interest. In simpler terms, it allows us to predict the



(a) `/r/PC`



(b) `/r/PCM`

Figure 5.1: Odds ratios between empirical and random conditional probabilities of interaction for `/r/PC` (a) and `/r/PCM` (b), with respect to social (left) and economic (right) axes. The interactions show a homophilic pattern on the social axis (higher values in the main diagonal) and a heterophilic one on the economic axis (higher values in the anti-diagonal).

likelihood of an event occurring based on a set of features. Logistic regression works by estimating how different factors (represented by independent variables, often denoted as $\mathbf{X} \in \mathbb{R}^{n \times m}$, where n is the size of the dataset and m is the number of features) affect a binary outcome variable (denoted as $y \in \{0, 1\}$). By analyzing a dataset containing both the features and the corresponding outcomes, the model estimates the relationships between these factors and predicts the probability of the outcome for new data points. We use this approach to predict the likelihood of user interactions within a social network (directed weighted interaction graph $G = (V, E, w)$ as defined in Section 4.1). By analyzing user features, we aim to estimate which combinations of features increase or decrease the probability of user interaction occurring within the network.

For a given node pair $u, v \in V$ the *target variable* is:

$$y_{u,v} = \begin{cases} 1, & \text{if } (u,v) \in E \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

We assume that the likelihood of observing an interaction $u \to v$ might be influenced by the combined feature values of both u and v. Each node $v \in V$ has $|F|$ demographic features, with $F := \{\text{young, old, male, female, poor, rich}\}$. We represent such characteristics as a vector $\mathbf{x}_u \in \{0,1\}^{|F|}$, with $x_{u,i} = 1$ if and only if the node u is in the quartile of considered users more likely to be in the demographic group i, according to Section 3.2.

Thus, the *independent variable* can be represented as:

$$\mathbf{X}_{u,v} = \mathbf{x}_u \otimes \mathbf{x}_v \in \{0,1\}^d \subseteq F \times F \tag{5.3}$$

where $d = |F|^2$, and $\mathbf{x}_u \otimes \mathbf{x}_v$ is the outer-product.

In this context, the logistic regression model is trained on a series of pairs $(y_{u,v}, \mathbf{X}_{u,v})$, producing as output coefficients a feature-feature matrix M. This estimated matrix provides insights into which demographic groups are more likely to interact. More explicitly, each matrix entry M_{ij} denotes the log-odds ratio of a node with feature value i to interact with a node with feature value j, compared to the probability of random interactions given by the null model.

To define the non-existing edges for the training (where $y_{u,v} = 0$), we employ the configuration model in Section 4.3. We balance positive and negative edge examples by shuffling the existing interaction network edges. We choose edges based on the source node u activity and the target node v attractiveness, according to Algorithm 1. If such a pair already exists, it is omitted. This method ensures that the estimated matrix M correctly represents how the demographic features make the observed edges deviate from this null model.

5.2.2 Homophily within Demographics Groups

Figure 5.2 shows the odds ratios, as exponentiated logistic regression coefficients, for each ordered pair of interacting demographic features on both `/r/PC` (a) and `/r/PCM` (b). Only coefficients significant at the $\alpha = 5\%$ level are shown.

Figure 5.2a reveals pronounced homophily among users based on their demographic features: pairs on the diagonal (i.e., within-class interactions) occur more frequently than expected. Especially interaction among 'old' users is 14% more likely than expected, and interaction among 'young' and 'female' users is 7-9% more likely.

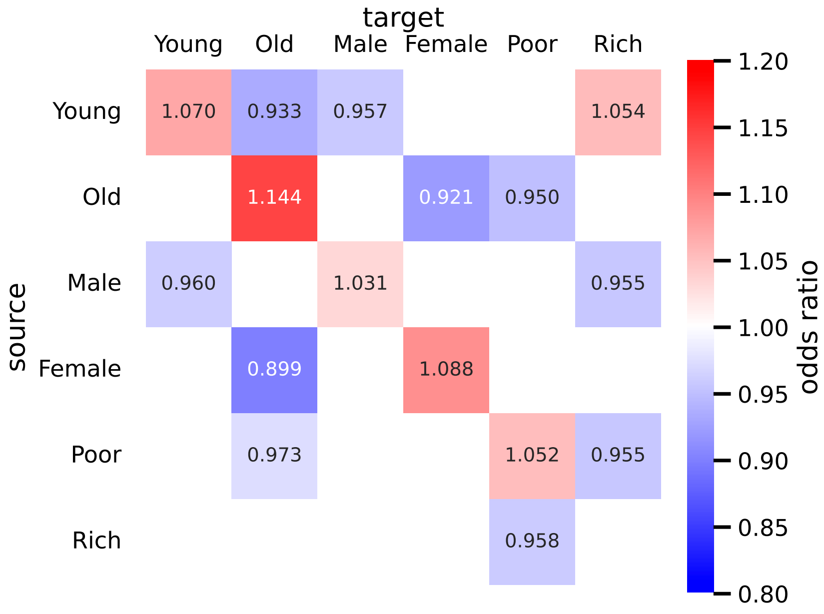
Figure 5.2b shows the same pronounced homophily for `/r/PCM`, especially interactions among 'young', 'old', and 'poor', which are respectively 15%, 13% and 9% more likely than expected.

Conversely, off-diagonal elements show a lower-than-expected interaction probability between users with different demographic features. In Figure 5.2b for `/r/PCM` this effect is particularly notable for the age feature, where 'old' and 'young' interactions are infrequent (odds ratios below 0.90).

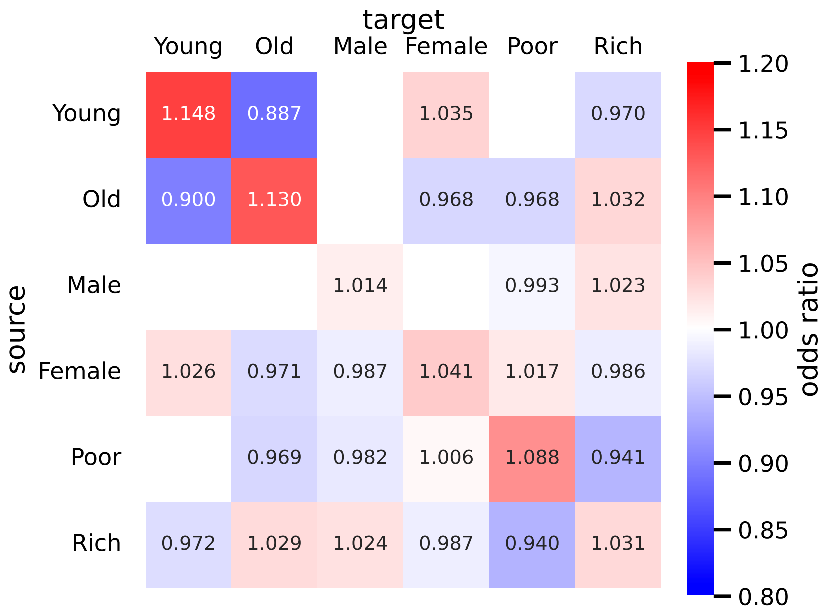
5.3 Modeling the Combined Effects: Ideology, Demographics, and Confounding Effects

Next, we analyze the interplay between ideological and demographic features and their joint effect on user interactions. Additionally, we recognize the potential influence of a user's popularity on Reddit (confounding), which could bias the interactions.

To ensure the significance of our findings, we utilize the logistic regression approach introduced in Section 5.2.1.



(a) `/r/PC`



(b) `/r/PCM`

Figure 5.2: Odds ratio (exponentiated logistic regression coefficients) for each ordered pair of interacting features on `/r/PC` (a) and `/r/PCM` (b). The source user is in the rows and the target user is in the columns. Only coefficients significant at the $\alpha = 5\%$ level are shown. The results show homophily in the demographic attributes with higher values in the main diagonal.

5.3.1 Logistic Regression Model for Features Importance

It is crucial to avoid the risk of multicollinearity to obtain the statistical significance of the interaction patterns for both the ideological and demographic dimensions of users employing the logistic regression model. *Multicollinearity* occurs in a regression model when independent

variables are highly correlated and can bias the results.
As logistic regression used in Section 5.2.1, all variables are represented as binary. Alongside demographic features, variables for two-dimensional political ideology should also be added. One approach might be to create 3 additional binary variables for each ideological dimension (social and economic): this would result in 6 total variables, as 'left', 'right', and 'economic center', and 'libertarian', 'authoritarian', and 'social center'. However, this approach can lead to multicollinearity. For instance, if a user identifies as having a 'left' economic ideology, then the variables 'economic center' and 'right' would be highly correlated with 'left' (as they cannot all be true simultaneously).
Aware of the risk of multicollinearity, we opt for a selected combination of ideological and demographic values rather than examining all pairwise combinations. As indicated by Equation (5.3), the independent dummy variables for our analysis are denoted as $\mathbf{X}_{u,v} \in \{0,1\}^{29}$. The 29 selected features include:

i **Ideological Features:** we consider the following set of pairwise features associated with each source-target edge ($u \to v$). Given $X_u, X_v \in \{\text{left}, \text{right}\}$, we define

- Economic Homophily $= 1$ if $X_u = X_v$
- Economic Heterophily $= 1$ if $X_u \neq X_v$

Given $X_u, X_v \in \{\text{libertarian}, \text{authoritarian}\}$, we define

- Social Homophily $= 1$ if $X_u = X_v$
- Social Heterophily $= 1$ if $X_u \neq X_v$

In both cases, users labeled as 'center' do not contribute to the features. We represent the directionality of the interaction with a separate feature denoted by an arrow. This design choice helps us to distinguish between homophilic/heterophilic effects and their asymmetry. As a result, we identify 6 possible ideological features, 2 economic features, 2 social features, and 2 features for asymmetry.

ii **Demographic Features:** We consider all the pairwise combinations of values of each feature (age: young-old, gender: male-female, affluence: poor-rich) as those employed to analyze the effect of demographics. Therefore, we have $6 \times (6+1)/2 = 21$ possible demographic features.

iii **Confounding Feature**: we take the popularity of a user introduced in Section 2.4.2.
We quantile-normalize the popularity values and define 4 classes of popularity. The top and bottom quartiles are considered as two distinct binary features. By taking into account the target's popularity, we define:

- Target is popular $= 1$ if the target is in the top quartile
- Target is not popular $= 1$ if it is in the bottom quartile

We thus add 2 confounding features for popularity.

Table 5.1 summarizes the full set of features we use in the logistic regression model to assess the interplay between ideological and demographic features and their joint effect on user interactions, in addition to the confounding effects of Reddit.
Figure 5.3 reports the results of the logistic regression model for both `/r/PC` and `/r/PCM`. A coefficient greater than 0 positively impacts the likelihood of the interaction, while a coefficient less than 0 has a negative impact. Each coefficient of the model is reported with a 99% confidence interval and its statistical significance.

Table 5.1: Set of features we use in the logistic regression model to model the combined effects of ideology, demographics, and confounding Reddit effects.

Feature Name	Feature Type		
	Ideological	Demographic	Confounding
Homophily (economic)	✓		
Heterophily (social)	✓		
Heterophily (Left →)	✓		
Heterophily (Lib →)	✓		
Homophily (social)	✓		
Heterophily (economic)	✓		
Young ↔ Old		✓	
Poor ↔ Rich		✓	
Old ↔ Female		✓	
Old ↔ Poor		✓	
Young ↔ Rich		✓	
Male ↔ Poor		✓	
Male ↔ Female		✓	
Female ↔ Rich		✓	
Young ↔ Male		✓	
Old ↔ Male		✓	
Female ↔ Poor		✓	
Young ↔ Poor		✓	
Male ↔ Male		✓	
Male ↔ Rich		✓	
Old ↔ Rich		✓	
Young ↔ Female		✓	
Rich ↔ Rich		✓	
Female ↔ Female		✓	
Poor ↔ Poor		✓	
Old ↔ Old		✓	
Young ↔ Young		✓	
Target is popular			✓
Target is not popular			✓

Economic Heterophily and Social Homophily

Figure 5.3 shows that these results statistically confirm previously observed trends in the influence of ideology on interactions (Figure 5.1). Having opposite economic ideologies increases the odds of interaction by 10% for `/r/PC` and more than 16% for `/r/PCM`. This effect is mirrored by a 5% decrease in the odds of interaction when both users have the same economic ideology for `/r/PC` and a 12% decrease for `/r/PCM`.

Conversely, pairs of users with the same social ideology are more likely to interact, with an increment of the odds of 8% for `/r/PC` and slightly more than 6% for `/r/PCM`. Similarly, heterophilic interactions across different social ideologies are less likely by 7% for `/r/PC` and slightly higher than 6% for `/r/PCM`.

The results for directional heterophily are not statistically significant at the $\alpha = 5\%$ level for `/r/PC`. However, for `/r/PCM`, it shows a weak influence on interactions.

Demographic Homophily

Figure 5.3 reinforces the initial findings presented in Figure 5.2, by providing statistically significant confirmation of the demographic effects on user interactions. Age homophily significantly

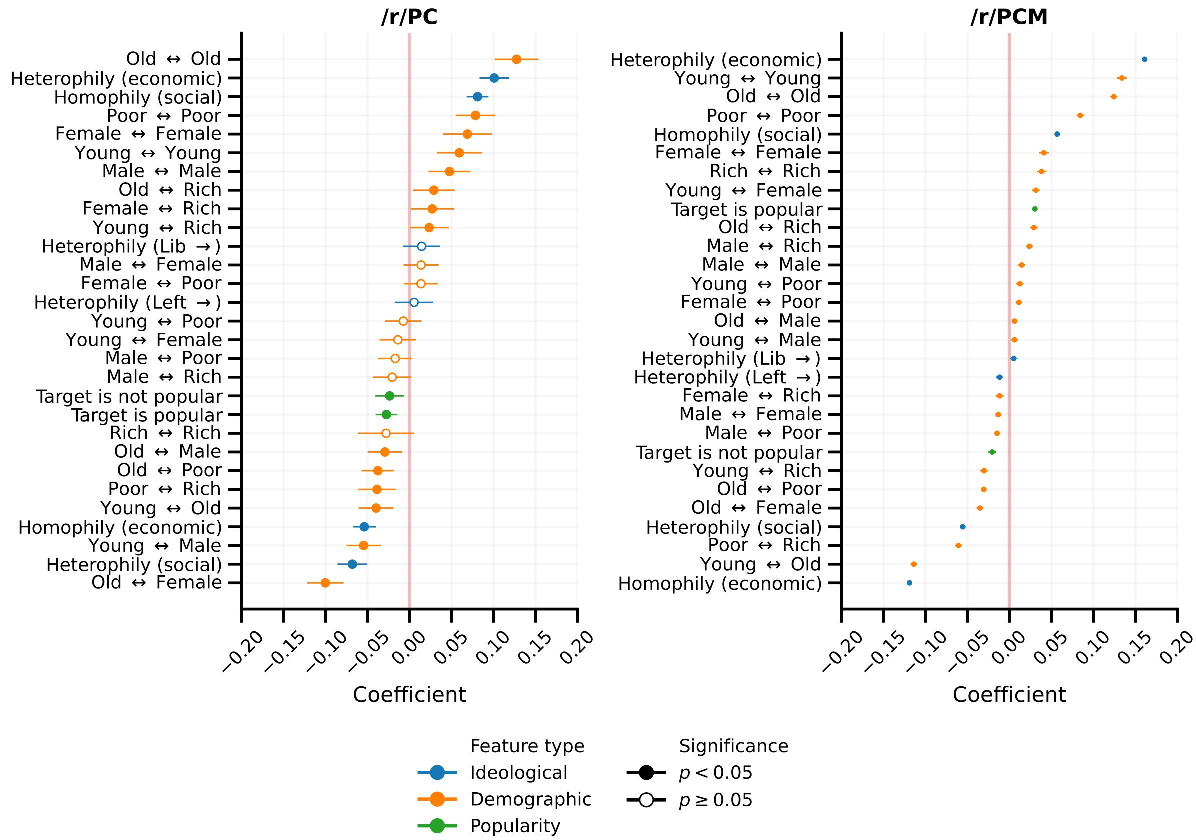


Figure 5.3: Coefficients and 99% confidence intervals for logistic regression features on both /r/PC and /r/PCM. The features are displayed in the rows and they represent pairs of classes in ideological characteristics (blue), demographic characteristics (orange), or popularity (green). A coefficient greater than 0 positively impacts the likelihood of interaction. Statistical significant results are highlighted with full markers.

predicts interactions, with a 13% increase in odds for ‘old’–‘old’ pairs and a 6% increase for ‘young’–‘young’ pairs in /r/PC. In /r/PCM, these odds increase by 13% and 14%, respectively. Demographic homophily in terms of ‘poor’–‘poor’, ‘female’–‘female’, and ‘male’–‘male’ interactions shows higher likelihood, with odds increasing by 8%, 7%, and 5% for /r/PC and 8%, 4%, and 2% for /r/PCM. Results for ‘rich’–‘rich’ interactions are not statistically significant at the $\alpha = 5\%$ level for /r/PC. However, in /r/PCM, they cause an odds increment of 4%.

Generally, heterophily in demographics reduces the odds of interactions. In particular, ‘poor’–‘rich’ and ‘young’–‘old’ interactions in /r/PC decrease odds both by 4%. In /r/PCM, these reductions are 6% and 12%.

Furthermore, cross-feature interactions generate different results. ‘Old’–‘poor’ and ‘old’–‘female’ interactions in /r/PC decrease odds by 4% and 10%. In /r/PCM, these reductions are both of 4%. ‘Old’–‘rich’ interactions cause a 3% increase in odds for both /r/PC and /r/PCM. ‘Female’–‘rich’ and ‘young’–‘rich’ interactions yield opposite results, with odds increasing in /r/PC and decreasing in /r/PCM. ‘Old’–‘male’ and ‘young’–‘male’ interactions lead to a decrease in odds for /r/PC but a small increase for /r/PCM. ‘Young’–‘female’ interactions are not statistically significant for /r/PC but result in an odds increment for /r/PCM.

5.4 Analyzing the Language Toxicity in Social Interaction

We found a higher rate of heterophilic interactions on the economic axis of the political compass. Here we show that heterophilic interactions are generally characterized by higher *toxicity*, which hints at conflictual interactions between axes poles. We use Google's Perspective API[1] to determine the toxicity scores of comments from both `/r/PC` and `/r/PCM`, excluding comments containing only emojis or links. It is a free tool for developers and online platforms, that utilizes machine learning to analyze text content. This functionality allows for predicting the perceived impact of a comment on a conversation, ultimately promoting healthier online discussions and content moderation. We apply this tool to the `body` attribute of collected comments (Section 2.3). In this context, **toxicity** is defined as "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion"[2].

A toxicity score τ ranges between 0 (lowest) and 1 (highest), and indicates the likelihood that an individual perceives the text as toxic. For each pair of interacting ideologies (X, Y, with $X \to Y$), $\overline{\tau}_{X \to Y}$ denotes the average toxicity of the comments exchanged. In Section 4.3, we used a null model to isolate the impact of network structure on various network properties. In this section, we employ a similar approach to determine whether the observed toxicity patterns within these interaction networks are a result of the network structure (interaction patterns) or simply due to chance.

However, the null model must meet two important criteria in this analysis. The first is to preserve the degree distribution of the nodes, similar to the previously used null model, i.e. to preserve the number of connections each user has in the original network. Second, unlike the previous null model, it must also preserve the original toxicity distribution of the interactions, referring to the toxicity scores of the comments exchanged within the network.

By incorporating these two properties, a null model can be created that closely resembles the real network in terms of structure and inherent toxicity levels within the interactions. This allows for a comparison between the observed toxicity patterns in the real network and the expected patterns in a random network with similar structural properties and toxicity levels. Significant deviations from the null model would indicate that the network structure itself influences the observed toxicity patterns.

5.4.1 Null Model for Toxicity Analysis

As shown in Table 4.1, the interaction networks for both `/r/PC` and `/r/PCM` contain a large interaction number W. Evaluating the toxicity of each comment within these networks presents a significant computational challenge. To overcome this issue, we selected a subset of interactions for analysis. For each network (`/r/PC` and `/r/PCM`), we extracted a sample of 100 000 comments. This sample size represents a balance between achieving statistical significance and ensuring a good representation of the overall toxicity distribution within the network. Furthermore, given the significant computational resources required for large-scale toxicity analysis, 100 000 comments represent a manageable workload that can be processed within a reasonable timeframe using our available computing architecture. This approach allows us to efficiently evaluate the toxicity of comments while maintaining a representative sample size for robust analysis. Therefore, we randomize only the political ideologies of the child and parent nodes within the network, while maintaining the original node degree and toxicity distributions.

To ensure a more robust analysis of the empirical network properties, we randomly sample 100 null models instead of relying on a single one, as we are analyzing a subset of 100,000 comments. To achieve this, we utilize a Monte Carlo Markov Chain (MCMC) edge-swapping technique [28]. The MCMC process operates iteratively by randomly selecting two target nodes and swapping

[1] https://perspectiveapi.com
[2] https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages

Algorithm 2: Sampling with edge-swapping.

Input: Edge list E of 100k sampled links; Switch threshold $\sigma = \log|E| \cdot |E|$.
Output: Edge list E with shuffled target nodes.

$s \leftarrow 0$;
while $s < \sigma$ **do**
  randomly select a target node v_i with the uniform probability with $(u_i, v_i) \in E$;
  randomly select a target node v_j with the uniform probability with $(u_j, v_j) \in E$;
  while $v_i = v_j$ ***or*** $u_i = v_j$ ***or*** $u_j = v_i$ **do**
    randomly select a target node v_j with the uniform probability with $(u_j, v_j) \in E$;
  $E \leftarrow E \cup \{(u_i, v_j), (u_j, v_i)\}$;
  $E \leftarrow E \setminus \{(u_i, v_i), (u_j, v_j)\}$;
  $s \leftarrow s + 1$;
return E

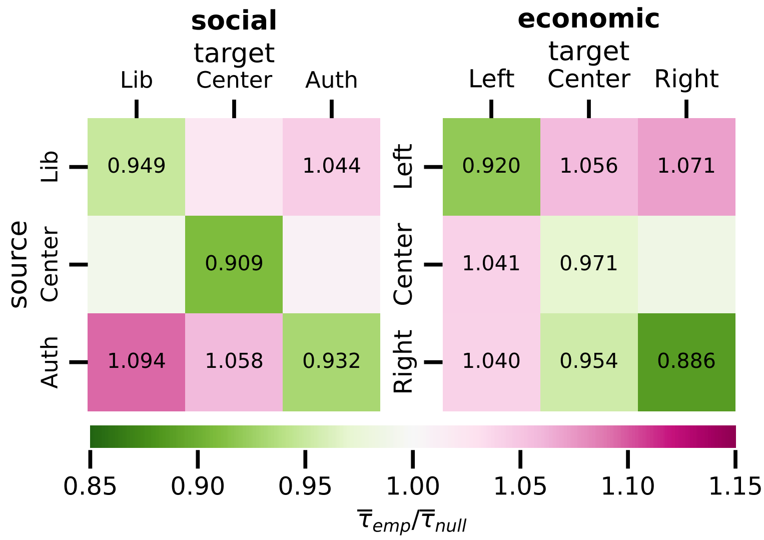
them, as long as no self-loop is created. This process is repeated for a total of $Q \cdot |E|$ edge-swappings, where $|E|$ is the edge count of the network. The technique essentially rewires the connections between users while preserving the overall nodes degree (in- and out-degree) and toxicity distributions of the network. By performing 100 MCMC edge-swapping simulations, we obtain 100 individual configurational models. To evaluate Markov chain convergence, we use a conservative value of $Q = \log|E|$, as recommended by Uzzi et al. [63].
Algorithm 2 shows our approach for sampling a configurational model using the MCMC edge-swapping technique. We perform $\sigma = \log|E| \cdot |E| = 500\,000$ edge swaps (given $|E| = 100\,000$ edges). The *acceptance rule* for this MCMC process is to prevent self-loops during the edge swapping. Finally, we combine all 100 configurational models into a single comprehensive null model. This null model represents an ensemble of random network structures with the same properties as the original network.

5.4.2 Quantifying Toxicity Levels of Interactions

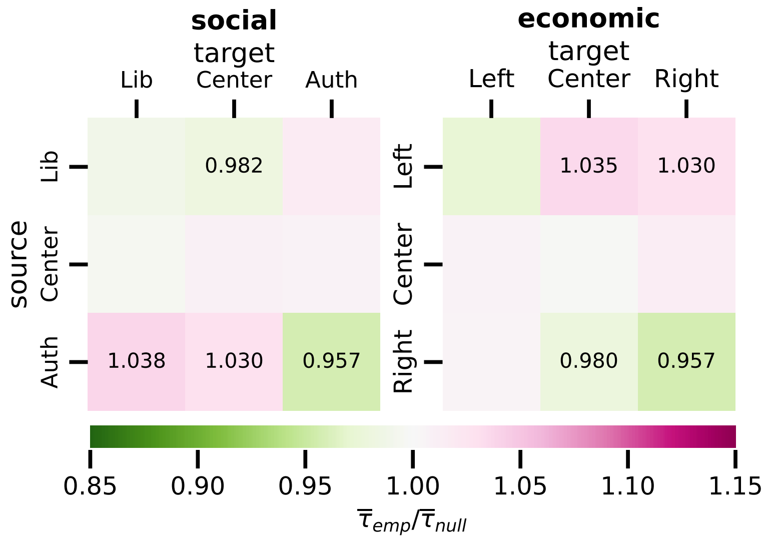
In this section, we quantify the toxicity levels within interactions between different ideological groups. We employ the comprehensive null model introduced in the previous section.
The null model allows us to estimate the *expected* level of toxicity for a given interaction $X \rightarrow Y$ between two ideologies (X, Y) in the network. We denote this expected toxicity as $\bar{\tau}^{\text{null}}_{X \rightarrow Y}$. Next, we calculate the actual average toxicity $\bar{\tau}^{\text{emp}}_{X \rightarrow Y}$ observed in the empirical interaction network for the same interaction $(X \rightarrow Y)$.
To compare these values, we compute the ratios:

$$\frac{\bar{\tau}^{\text{emp}}_{X \rightarrow Y}}{\bar{\tau}^{\text{null}}_{X \rightarrow Y}}.$$

Ratios greater than 1 indicate that the empirical toxicity between ideologies X and Y is higher than what would be expected by chance in a random network with similar properties (as captured by the null model).
Following this initial assessment, we perform a t-test to evaluate the statistical significance of these observed differences. This statistical test helps determine if the higher toxicity levels observed in the real network are statistically meaningful or simply due to random fluctuations. We set a significance level of $\alpha = 5\%$. Results with p-values exceeding this threshold (non-significant) are excluded from analysis, indicating that the observed difference in toxicity might be due to chance. In essence, this approach allows us to systematically quantify toxicity



(a) /r/PC



(b) /r/PCM

Figure 5.4: Ratio between the average toxicity in empirical data ($\overline{\tau}_{\text{emp}}$) and random null model ($\overline{\tau}_{\text{null}}$) for /r/PC. Values above 1 indicate higher toxicity than expected. Non-significant results at the $\alpha = 5\%$ level are omitted (via a t-test). The left heatmap shows interactions between ideologies on the social axis, and the right heatmap focuses on the economic axis. Heterophilic interactions (lib-auth, left-right) exhibit higher toxicity than expected.

levels within ideological interactions in the network and assess whether these levels deviate significantly from what would be expected in a random scenario.

Figure 5.4a depicts the ratio of the average empirical toxicity to the average toxicity in the null model for /r/PC. On both axes, heterophilic comments present a higher toxicity than expected, while homophilic comments present a lower one (on-diagonal values are lower than 1, while off-diagonal ones are higher than 1). On the economic axis, heterophilic comments (i.e., between left- and right-leaning users) are 4-to-7% more toxic than expected, and statistically significant in both directions (p-values of 0.0004 and $< 10^{-6}$). Conversely, interactions between users of the same ideology demonstrate lower-than-expected toxicity with an effect of 8-to-11% (p-values $< 10^{-6}$).

Regarding the social axis, homophilic comments between users of identical political ideologies are 5-to-9% less toxic than expected (p-values $< 10^{-6}$). Instead, comments from authoritarians towards libertarians are significantly more toxic (10%, p-value $< 10^{-6}$), with a similar albeit smaller effect in the opposite direction.

People who interact with others who have different political views (Figure 5.4a) are more likely to engage in toxic behavior, while people who interact with others who have the same political views are less likely to engage in toxic behavior.

The same considerations apply to `/r/PCM`, but the results are less evident and/or not statistically significant, as shown in Figure 5.4b. This suggests that a more humorous discussion involving memes, as those in `/r/PCM`, may be less conflictual than a serious discussion, such as those in `/r/PC`.

Chapter 6

Conclusion

In our quantitative analysis of interaction patterns on Reddit's communities /r/PoliticalCompass and /r/PoliticalCompassMemes from 2020 to 2022, we uncovered dynamics that go beyond the conventional left–right political dichotomy. Users with similar social ideologies and demographics interacted more frequently and typically used non-toxic language, which denotes a high level of homophily on the social axis of the political compass. Instead, conversations reflecting "social" heterophily, particularly between authoritarian and libertarian ideologies, occurred 6% less frequently than expected. In contrast, heterophilic and conflictual interactions were pronounced on the economic axis, with users of opposing economic ideologies displaying higher signs of toxicity and engaging with each other 10% more than expected.

In light of our results, some of the apparently puzzling differences in the recent literature may be reconciled. On the one hand, social media platforms such as Twitter and Facebook, which emphasize social connections, reportedly exhibit a high level of homophily [19]. The presence of political echo chambers [15, 31, 2] is therefore to be expected if this social homophily is a driver of the interactions, as suggested by our results. On the other hand, platforms such as Reddit, where status homophily is less dominant and interactions are centered on shared topical interests, tend to experience more conflictual interactions across ideologies [62]. This fact is particularly true for Reddit's political spaces, which skew towards the U.S., a nation where the political spectrum largely corresponds to a singular economic left-right dimension [39]. As a result, echo chambers may be rarer on platforms like Reddit [15, 20, 49]. Building on our findings, it is evident that the nature of interactions, in terms of toxicity along with the specific ideological axis, significantly influences the manifestation of homophily and heterophily.

Our findings also highlight the inclination of users to group based on demographics and similar social ideologies, hinting at a certain degree of social affinity [64]. Belonging to such digital echo chambers provides a sense of community but also poses risks. Being predominantly exposed to only one type of ideology can fuel misinformation [61, 23]. It can strengthen existing biases and further spread incorrect beliefs. If these online tendencies continue, they might intensify real-world divisions with consequences in voting patterns and everyday interactions [54].

Our study is not exempt from limitations. Firstly, our dataset lacks geographical granularity; it focuses on English-speaking users predominantly from a U.S.-centric platform, Reddit. Future studies should aim to infer and incorporate the geographical locations of users to control for regional effects and limit sampling bias. Extracting self-declarations from plain text or inferring them via machine learning models could extend the dataset to additional subreddits and social networks, thus improving result generalizability. Secondly, our reliance on self-declarations means that we excluded users who change their political orientation in time. An interesting direction for future research is analyzing such opinion changes, under the lens of the political discussions that occurred within the subreddit and outside of it.

Moreover, user declarations may not strictly align with the true ideology of users nor with the political content in the observed subreddits. While declared ideologies along the economic axis are validated by comparison to left/right ideologies inferred by following Waller and Anderson [65], we could not validate self-declarations along the social axis. Future work could be devoted to building a more comprehensive embedding of Reddit users that takes into account their political stance in a multidimensional way, e.g., by adding the libertarian–authoritarian axis.

Likewise, the textual content of the interactions is a rich resource that can yield further insights and has not yet been fully utilized.

Lastly, our findings pave the way for multidimensional modeling and intervention studies. For instance, a social compass model has been recently proposed to explore depolarization dynamics in multidimensional topics represented in a polar space [50, 51]. Similarly interesting directions for future work are the exploration of algorithms that offer diverse content by taking into account the multidimensional nature of targeted users [30, 33].

Bibliography

[1] Amy Adamczyk. "Religion as a Micro and Macro Property: Investigating the Multilevel Relationship between Religion and Abortion Attitudes across the Globe". In: *European Sociological Review* (Mar. 2022), jcac017. ISSN: 0266-7215. DOI: 10.1093/esr/jcac017.

[2] Jisun An, Daniele Quercia, and Jon Crowcroft. "Partisan Sharing: Facebook Evidence and Societal Consequences". In: *COSN'14: ACM Conference on Online Social Networks*. 2014, pp. 13–24.

[3] Chris Bail. *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton University Press, Dec. 2021. ISBN: 978-0-691-21650-8. DOI: 10.1515/9780691216508. URL: https://www.degruyter.com/document/doi/10.1515/9780691216508/html.

[4] Delia Baldassarri and Andrew Gelman. "Partisans without Constraint: Political Polarization and Trends in American Public Opinion". In: *AJS* 114.2 (Jan. 2008), pp. 408–446. ISSN: 0002-9602. DOI: 10.2139/ssrn.1010098.

[5] Michele Ballerini et al. "Empirical investigation of starling flocks: a benchmark study in collective animal behaviour". In: *Animal Behaviour* 76.1 (2008), pp. 201–215. ISSN: 0003-3472. DOI: https://doi.org/10.1016/j.anbehav.2008.02.004. URL: https://www.sciencedirect.com/science/article/pii/S0003347208001176.

[6] Pablo Barberá. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data". In: *Political analysis* 23.1 (2015), pp. 76–91.

[7] Michael Barthel et al. "Reddit news users more likely to be male, young and digital in their news preferences". In: *Pew Research Center* (2016).

[8] Fabian Baumann et al. "Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces". In: *Phys. Rev. X* 11.1 (Jan. 2021), p. 011012. DOI: 10.1103/PhysRevX.11.011012.

[9] Fabian Baumann et al. "Modeling Echo Chambers and Polarization Dynamics in Social Networks". In: *Phys. Rev. Lett.* 124 (4 2020), p. 048301. DOI: 10.1103/PhysRevLett.124.048301.

[10] Jason Baumgartner et al. "The pushshift reddit dataset". In: *arXiv preprint arXiv:2001.08435* (2020).

[11] Kenneth Benoit and Michael Laver. "The dimensionality of political space: Epistemological and methodological considerations". en. In: *European Union Politics* 13.2 (June 2012), pp. 194–218. ISSN: 1465-1165. DOI: 10.1177/1465116511434618.

[12] A. Bessi et al. "Users polarization on Facebook and Youtube". In: *PLOS ONE* 11.8 (2016), e0159641.

[13] Max Bradley and Simon Chauchard. "The Ethnic Origins of Affective Polarization: Statistical Evidence from Cross-National Data". In: *Frontiers in Political Science* (2022), p. 64.

[14] Tinggui Chen et al. "Modeling Multidimensional Public Opinion Polarization Process under the Context of Derived Topics". en. In: *International Journal of Environmental Research and Public Health* 18.2 (Jan. 2021), p. 472. ISSN: 1660-4601. DOI: 10.3390/ijerph18020472.

[15] Matteo Cinelli et al. "The Echo Chamber Effect on Social Media". In: *PNAS: Proceedings of the National Academy of Sciences* 118.9 (2021), e2023301118. DOI: 10.1073/pnas.2023301118. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.2023301118.

[16] Michael D Conover et al. "Political polarization on twitter". In: *Fifth international AAAI conference on weblogs and social media*. 2011.

[17] Wesley Cota et al. "Quantifying echo chamber effects in information spreading over political communication networks". In: *EPJ Data Science* 8.1 (2019), p. 35. DOI: 10.1140/epjds/s13688-019-0213-9. URL: https://doi.org/10.1140/epjds/s13688-019-0213-9.

[18] Chad Crawford, Logan Brooks, and Sandip Sen. "Opposites repel: the effect of incorporating repulsion on opinion dynamics in the bounded confidence model". In: *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. AAMAS '13. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, May 2013, pp. 1225–1226. ISBN: 978-1-4503-1993-5.

[19] Munmun De Choudhury. "Tie formation on twitter: Homophily and structure of egocentric networks". In: *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE. 2011, pp. 465–470.

[20] Gianmarco De Francisci Morales, Corrado Monti, and Michele Starnini. "No Echo in the Chambers of Political Interactions on Reddit". In: *Scientific Reports* 11 (2021), p. 2818. DOI: 10.1038/s41598-021-81531-x.

[21] Guillaume Deffuant et al. "Mixing beliefs among interacting agents". In: *Advs. Complex Syst.* 03.01n04 (Jan. 2000), pp. 87–98. ISSN: 0219-5259. DOI: 10.1142/S0219525900000078.

[22] Michela Del Vicario et al. "Echo chambers: Emotional contagion and group polarization on facebook". In: *Scientific reports* 6 (2016). Publisher: Nature Publishing Group, p. 37825.

[23] Michela Del Vicario et al. "Echo chambers in the age of misinformation". In: *arXiv preprint arXiv:1509.00189* (2015).

[24] Daniel DellaPosta, Yongren Shi, and Michael Macy. "Why Do Liberals Drink Lattes?" In: *American Journal of Sociology* 120.5 (Mar. 2015), pp. 1473–1511. ISSN: 0002-9602. DOI: 10.1086/681254.

[25] Elizabeth Dubois and Grant Blank. "The echo chamber is overstated: the moderating effect of political interest and diverse media". In: *Information, communication & society* 21.5 (2018), pp. 729–745.

[26] John V Duca and Jason L Saving. "Income inequality and political polarization: time series evidence over nine decades". In: *Review of Income and Wealth* 62.3 (2016), pp. 445–466.

[27] Fabian Falck et al. "Measuring Proximity Between Newspapers and Political Parties: The Sentiment Political Compass". en. In: *Policy & Internet* 12.3 (2020), pp. 367–399. ISSN: 1944-2866. DOI: 10.1002/poi3.222.

[28] Bailey K. Fosdick et al. *Configuring Random Graph Models with Fixed Degree Sequences*. Tech. rep. arXiv:1608.00607. arXiv:1608.00607 [physics, q-bio, stat] type: article. arXiv, Oct. 2017. URL: http://arxiv.org/abs/1608.00607.

[29] André Freire. "Party Polarization and Citizens' Left—Right Orientations". en. In: *Party Politics* 14.2 (Mar. 2008), pp. 189–209. ISSN: 1354-0688. DOI: 10.1177/1354068807085889.

[30] Kiran Garimella et al. "Factors in Recommending Contrarian Content on Social Media". In: *9th International ACM Web Science Conference*. WebSci. 2017, pp. 263–266.

[31] Kiran Garimella et al. "Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship". In: *The Web Conference*. WWW. 2018, pp. 913–922.

[32] Kiran Garimella et al. "Quantifying Controversy in Social Media". In: *WSDM '16: 9th ACM International Conference on Web Search and Data Mining*. 2016, pp. 33–42.

[33] Kiran Garimella et al. "Reducing Controversy by Connecting Opposing Views". In: *ACM International Conference on Web Search and Data Mining*. WSDM. [best student paper award], 2017, pp. 81–90.

[34] R Kelly Garrett. "Echo chambers online?: Politically motivated selective exposure among Internet news users". In: *Journal of Computer-Mediated Communication* 14.2 (2009), pp. 265–285.

[35] Eric Gilbert, Tony Bergstrom, and Karrie Karahalios. "Blogs are echo chambers: Blogs are echo chambers". In: *42nd Hawaii International Conference on System Sciences*. 2009, pp. 1–10.

[36] Andrew Guess et al. "Avoiding the echo chamber about echo chambers". In: *Knight Foundation* (2018).

[37] Rainer Hegselmann and Ulrich Krause. "Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation". en. In: *Journal of Artificial Societies and Social Simulation* 5.3 (2002), pp. 1–2.

[38] A. Heywood. *Political Ideologies: An Introduction*. Macmillan education. Macmillan Education UK, 2017, pp. 14–17. ISBN: 9781137606013.

[39] Samara Klar. "A multidimensional study of ideological preferences and priorities among the American public". In: *Public Opinion Quarterly* 78.S1 (2014), pp. 344–359.

[40] Geoffrey C Layman, Thomas M Carsey, and Juliana Menasce Horowitz. "Party polarization in American politics: Characteristics, causes, and consequences". In: *Annual Review of Political Science, Vol 13* 9 (2006), pp. 83–110.

[41] Jan Clifford Lester. "Liberty and the political compass (or how left-wingism is anti-liberty)". In: *Journal of Social and Evolutionary Systems* 18.3 (1995), pp. 213–216.

[42] Jan Clifford Lester. "The political compass and why libertarianism is not right-wing". In: *Journal of social philosophy* 27.2 (1996), pp. 176–86.

[43] J.C. Lester. "The evolution of the political compass (and why libertarianism is not right-wing)". en. In: *Journal of Social and Evolutionary Systems* 17.3 (Jan. 1994), pp. 231–241. ISSN: 10617361. DOI: 10.1016/1061-7361(94)90011-6. URL: https://linkinghub.elsevier.com/retrieve/pii/1061736194900116.

[44] Lilliana Mason and Julie Wronski. "One tribe to bind them all: How our social group attachments strengthen partisanship". In: *Political Psychology* 39 (2018), pp. 257–277.

[45] Jennifer McCoy, Tahmina Rahman, and Murat Somer. "Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities". en. In: *American Behavioral Scientist* 62.1 (Jan. 2018), pp. 16–42. ISSN: 0002-7642. DOI: 10.1177/0002764218759576.

[46] Yelena Mejova et al. "Modeling Political Activism around Gun Debate via Social Media". en. In: *ACM Transactions on Social Computing*. TSC (Apr. 2022), p. 3532102. ISSN: 2469-7818, 2469-7826. DOI: 10.1145/3532102. URL: https://dl.acm.org/doi/10.1145/3532102.

[47] Lillio Mok, Michael Inzlicht, and Ashton Anderson. "Echo Tunnels: Polarized News Sharing Online Runs Narrow but Deep". en. In: *Proceedings of the International AAAI Conference on Web and Social Media* 17 (June 2023), pp. 662–673. ISSN: 2334-0770. DOI: 10.1609/icwsm.v17i1.22177. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/22177.

[48] Michael Molloy and Bruce Reed. "A critical point for random graphs with a given degree sequence". In: *Random structures & algorithms* 6.2-3 (1995), pp. 161–180.

[49] Corrado Monti et al. “Evidence of Demographic rather than Ideological Segregation in News Discussion on Reddit”. en. In: *Proceedings of the ACM Web Conference 2023*. Austin TX USA: ACM, Apr. 2023, pp. 2777–2786. ISBN: 978-1-4503-9416-1. DOI: 10.1145/3543507.3583468. URL: https://dl.acm.org/doi/10.1145/3543507.3583468.

[50] Jaume Ojer, Michele Starnini, and Romualdo Pastor-Satorras. “Modeling Explosive Opinion Depolarization in Interdependent Topics”. In: *Phys. Rev. Lett.* 130 (20 2023), p. 207401. DOI: 10.1103/PhysRevLett.130.207401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.207401.

[51] Jaume Ojer, Michele Starnini, and Romualdo Pastor-Satorras. *Vanishing threshold in depolarization of correlated opinions on social networks*. 2023. arXiv: 2306.01329 [physics.soc-ph].

[52] Andreas Petrik. “Core concept “Political Compass”. How Kitschelt’s model of liberal, socialist, libertarian and conservative orientations can fill the ideology gap in civic education”. In: *JSSE-Journal of Social Science Education* (2010).

[53] Joseph Phillips. “Affective Polarization: Over Time, Through the Generations, and During the Lifespan”. In: *Political Behavior* (2022), pp. 1–26.

[54] Samuel C Rhodes. “Filter bubbles, echo chambers, and fake news: how social media conditions individuals to be less critical of political misinformation”. In: *Political Communication* 39.1 (2022), pp. 1–22.

[55] Cesare Rollo et al. “Communities, Gateways, and Bridges: Measuring Attention Flow in the Reddit Political Sphere”. In: *Social Informatics: 13th International Conference, SocInfo 2022, Glasgow, UK, October 19–21, 2022, Proceedings*. Springer. 2022, pp. 3–19.

[56] Ana Lucía Schmidt et al. “Polarization of the vaccination debate on Facebook”. In: *Vaccine* 36.25 (2018), pp. 3606–3612.

[57] Simon Schweighofer, David Garcia, and Frank Schweitzer. “An agent-based model of multi-dimensional opinion dynamics and opinion alignment”. In: *Chaos* 30.9 (Sept. 2020), p. 093139. ISSN: 1054-1500. DOI: 10.1063/5.0007523.

[58] Alexander J. Stewart, Nolan McCarty, and Joanna J. Bryson. “Polarization under rising inequality and economic decline”. en. In: *Science Advances* 6.50 (Dec. 2020), eabd4201. ISSN: 2375-2548. DOI: 10.1126/sciadv.abd4201. URL: https://www.science.org/doi/10.1126/sciadv.abd4201.

[59] Michael Storper. “Separate Worlds? Explaining the Current Wave of Regional Economic Polarization”. en. In: *Journal of Economic Geography* 18.2 (Mar. 2018), pp. 247–270. ISSN: 1468-2702. DOI: 10.1093/jeg/lby011.

[60] John Toner and Yuhai Tu. “Long-Range Order in a Two-Dimensional Dynamical XY Model: How Birds Fly Together”. In: *Phys. Rev. Lett.* 75 (23 1995), pp. 4326–4329. DOI: 10.1103/PhysRevLett.75.4326. URL: https://link.aps.org/doi/10.1103/PhysRevLett.75.4326.

[61] Petter Törnberg. “Echo chambers and viral misinformation: Modeling fake news as complex contagion”. In: *PLoS one* 13.9 (2018), e0203958.

[62] Petter Törnberg. “How digital media drive affective polarization through partisan sorting”. en. In: *Proceedings of the National Academy of Sciences* 119.42 (Oct. 2022), e2207159119. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2207159119. URL: https://pnas.org/doi/full/10.1073/pnas.2207159119.

[63] Brian Uzzi et al. “Atypical combinations and scientific impact”. In: *Science* 342.6157 (2013), pp. 468–472.

[64] James A Vela-McConnell. *Who is my neighbor? Social affinity in a modern world*. SUNY Press, 1999.

[65] Isaac Waller and Ashton Anderson. "Quantifying social organization and political polarization in online platforms". In: *Nature* 600.7888 (2021), pp. 264–268.

[66] Ingmar Weber, Venkata R. Kiran Garimella, and Alaa Batayneh. "Secular vs. Islamist polarization in egypt on twitter". In: *2013 IEEE/ACM international conference on advances in social networks analysis and mining*. ASONAM. 2013, pp. 290–297. ISBN: 978-1-4503-2240-9. DOI: 10.1145/2492517.2492557. URL: http://dl.acm.org/citation.cfm?doid=2492517.2492557.