Class of 2024

PhD in Data Science

36th cycle

The Signed Ego Network: Modelling and Analysis Through the Lenses of Online Social Networks

Settore Scientifico Disciplinare: INF/01

Candidate
Jack David Tacchi

Supervisors
Chiara Boldrini
Andrea Passarella
Marco Conti

Academic year 2020–2024

ACKNOWLEDGMENTS

This thesis would not have been possible without the guidance and support of many individuals, to whom I am profoundly grateful. First, to Chiara Boldrini and Andrea Passarella for their enormous support throughout my PhD, always going far above and beyond. I have heard many people say that your supervisors are the most important part of your PhD and I did not realise until I started just how true that is. I couldn't have hoped for better supervisors and I would not have been able to finish without them, especially as they saved me from shooting myself in the foot in almost every paper I wrote.

I would also like to thank Marco Cont. He was the first person I met at CNR and took the time to introduce me to the institute and its labyrinthine halls. Marco always gave insightful and rigorous feedback that made sure I had an extensive grasp of the subject matter at hand.

I would also like to thank all of my friends who have also supported me throughout the highs and inevitable lows of the PhD. To Jesse, Nitin, Chris, Michele, Greco, Winston, Viktor, Andreas, Maria, Sam, Mustafa, Andrew, Kilian, Mattia, Robert, Emily, Abhishek, Marta, Lucy, Lorenzo, Mirko, Filippo, Jose, Josh and Ferdous. A life without friends is not truly lived.

I especially owe a huge debt of gratitude to my partner, Clelia, who has supported me since the day we met, despite my sense of humour. She has taught me enumerable things about both myself and life in general, showing me that, no matter how convinced I am of something, there is always room for another perspective.

Finally, I would like to thank my family, Charlotte, Jimmy, Alex, Lorraine, Scott and Sarah, as well as the two young additions of Freddie and Elowen. But most of all I am thankful to my parents, who have undoubtedly had the greatest impact on my life. Even though they do not always agree with my decisions, their unwavering love and support have allowed me to continue to grow as a person and for that, I am forever in their debt.

ABSTRACT

Humans have the ability to communicate at a scale and complexity that is unmatched by any other species on our planet. This capacity has been a key factor in allowing us to develop large-scale societies that are predominant across many diverse areas of the globe. What's more, since the advent of the internet, individuals and groups can be connected regardless of physical location. This increased connectivity has brought with it many new, emergent phenomena, not all of which are beneficial. Therefore, it is becoming increasingly important to understand the ways in which humans communicate and the social structures that underpin these interactive behaviours. One particularly pertinent model that facilitates this understanding is the Ego Network Model (ENM), which views a social network from the point-of-view of a single individual, organising their connections into circles around them based on the strength of their relationship. However, this model has the notable limitation of only measuring relationships based on the contact frequency of the individuals involved.

This thesis aims to establish an extension to the ENM that incorporates signed connections: the Signed Ego Network Model (SENM). To this end, a novel methodology for computing polarity sign (i.e. positive or negative) for the connections within a social network, based on sentiments of individual interactions, is first proposed. This method is shown to achieve similar results regardless of the model used to compute the individual sentiments and is also validated using known expectations of signed networks. Next, the signing methodology is used to compute the SENM, which is then investigated. The results of this reveal that, surprisingly, negativity is often most prevalent in the relationships we engage the most with. The potential effects of negativity on cognitive load are also investigated, although little statistically significant evidence was found. The ENM and SENM are then leveraged for the task of Stance Detection (SD), where they are able to be used to obtain results similar (although slightly worse) to the cutting edge, while using far less, and more easily obtainable, data. Finally, differences in the SENM are observed between cultures and online communities. Both cultures and engagement in a subcommunity were found to have an effect on the rate of negative relationships, although the former appears to be more influential. This is followed up by analyses of the most popular terms and talking points, which find that individuals who engage in more generic exchanges (e.g. about the weather) are more likely to have fewer negative relationships than those who commonly engage in more polarising topics (such as politics).

CONTENTS

1

INTRODUCTION

Humans are social animals. Every day we interact with other people and these interactions link us to each other to an extent that is not immediately obvious without observing the social network as a whole. Since the advent of the internet, people have been able to interact with each other with far greater ease, regardless of geographical location. In recent years and with the rise of online social networks (OSNs), communications have been facilitated even further. Indeed, it is now possible to instantly interact with someone on the other side of the globe with the mere click of a button. These technological advances have brought with them many advantages, including access to information at a scale that would have been unimaginable even a couple of generations ago. However, the swiftness with which they have been implemented means that many aspects of such global social networks, including many of their downsides, are not fully understood. Examples of drawbacks of global interconnectivity include biased exposure to information (echo chambers) (Cinelli et al., 2021; Nguyen, 2020) and long-distance radicalisation (Kadivar, 2017) but perhaps the most well-known is the prolific spread of misinformation, which has an increasingly influential impact on global events, most notably on elections (Munger et al., 2022; Swire et al., 2017).

Many of these phenomena are known to originate from the microscopic level features of a social network, which then cause emergent patterns and behaviours across the network at a much larger scale (i.e. at the meso- and macroscopic levels). For instance, information diffusion is often closely linked to the strength of individual relationships, with certain information only passing through relationships above a certain strength (Sutcliffe et al., 2012). Viewing this effect across a larger swathe of the network, one can often observe "highways" of information, where many strong connections allow for a free flow of information, and isolated groups surrounded by weak connections, where more "trusted" information is unable to penetrate. Given the importance of these microscopic structures, they have been studied quite extensively in the past, also with respect to Online Social Networks (OSNs). In this thesis, we aim to extend such models with an additional dimension: the *polarity* of social relationships. Specifically, we aim to derive a model, starting from the well-established Ego Network Model (ENM) presented in Section 1.2, that also captures the positive or negative signs of individual social relationships. Then, we aim to characterise such microscopic structures from the perspective of this polarity, to understand how polarity is distributed across the

various components of the ENM. To achieve this goal, we consider a significant body of work (summarised in the remainder of this Chapter) related to both the modelling of human social relationships from anthropology and psychology and to prior findings on how these models apply to OSNs.

Fortunately, significant efforts have been conducted towards improving our understanding of how humans communicate and, therefore, towards minimising the negative impacts of social media networks. Of course, human interactions existed long before the internet and they have a long history of study in many fields of research, which can provide excellent insights for dealing with the vast amounts of social data we now have access to. Anthropology in particular provides many more traditional theories and models that can be brought to bear against these modern social problems.

1.1 SOCIAL BRAIN HYPOTHESIS

One social pattern that has a particularly extensive history of observations is Dunbar's Social Brain Hypothesis (Dunbar, 1998). This hypothesis poses that the size of social group that a species can maintain is directly proportional to the neocortex region of its brain. If a social group exceeds the innate cognitive limit of its members then it will become unstable and start to fragment into smaller, more manageable groups (Dunbar, 1992). For humans, the number of active social connections that can be maintained is around 150, also known as Dunbar's number. Larger social groups can be seen to organise around this unit of group size throughout human history, from Neolithic village populations in Mesopotamia (Oates, 1977) to the sizes of independent units in professional armies of the 16^{th} to 20^{th} centuries (Dunbar, 1993), and even in modern-day communities in online social media (Arnaboldi et al., 2015). In fact, this inbuilt cognitive limit is so prevalent that it can also be observed in many species of non-human primates, albeit with smaller group sizes (Dunbar, 1998). An observable relationship between neocortex size and mean observed group size for various primates is displayed in Figure 1.

Of course, 150 is not the total number of individuals the average human interacts with throughout their lifetime but the number of actively maintained relationships at any single point in time. Research investigating human relationships based on the exchanging of Christmas cards found that relationships within the active network usually had a minimum interaction frequency of once per year (Hill and Dunbar, 2003). However, not all relationships within the 150 are equal. Despite the minimum frequency of interactions being annual, many relationships will have interactions monthly or even weekly. Therefore, it can be useful to categorise the relationships further into subgroups,

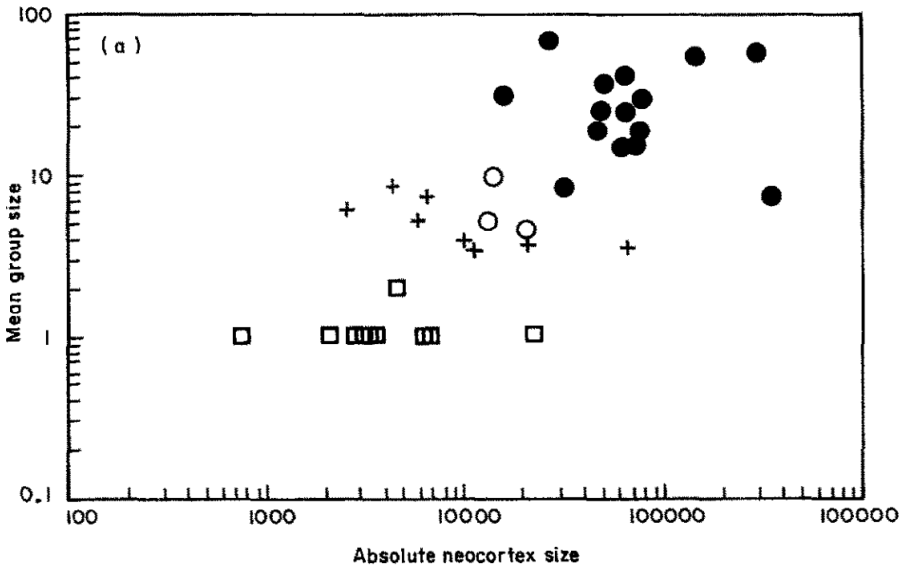

Figure 1: Relationship between neocortex size and observed mean group size for various primates, grouped into polygamous anthropoids (filled circles), monogamous anthropoids (pluses), diurnal prosimians (empty circles), nocturnal prosimians (squares) and hominoids (triangles). Taken from Dunbar, 1992.

in order to better understand them. One model which aims to do this is the Ego Network Model (ENM).

1.2 THE EGO NETWORK MODEL

The ENM centres around a single individual within a network, known as the *Ego*, from which the model takes its name. It then clusters all of their immediate connections, named *Alters*, around them based on the strength of their relationship. Doing this almost inevitably results in a series of concentric circles with increasing size but decreasing intimacy, as illustrated in Figure 2. For humans, the expected sizes of these circles range from 5 (support clique), to 15 (sympathy group), then around 45-50 (affinity group) and finally 150 (active network) (Dunbar and Spoors, 1995). Although one might assume that the ease of online communication would require less cognitive effort and therefore allow for larger social networks to be maintained, the ENM structure remains largely consistent in online contexts. Thus, it seems that no matter how advanced our tools of communication become, our social structures still appear to be limited by our innate cognitive limits (Miritello et al., 2013). Indeed, the only notable difference is the occasional presence of an additional innermost circle, with an average size of around 1.5 Alters (Arnaboldi et al., 2013; Dunbar et al., 2015). However, this additional circle has been postulated for offline networks as well, although quantities of offline data sufficient enough to confirm its existence have never been collected. The size ratio between each circle

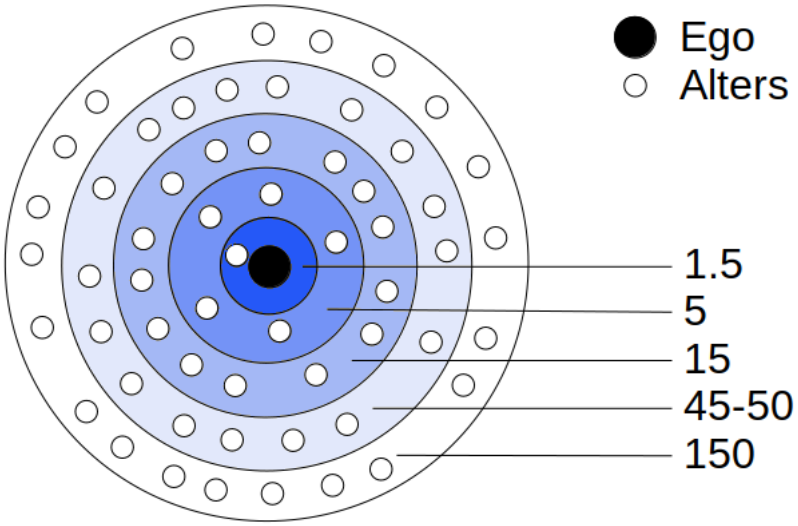

Figure 2: The Ego Network Model with the expected sizes of each circle.

is also notably consistent, with each subsequent circle increasing in size by a factor of around 3 (Hill and Dunbar, 2003).

Because each individual in a social network can be viewed as an Ego, the entire network itself can be thought of as a collection of interconnected Ego Networks. Thus, observing a network from the perspectives of the individual Egos can reveal insights that are only visible at a microscopic scale, yet have far-reaching consequences across the entire network. Indeed, the structural properties of the ENM have been shown to influence a number of social behaviours, such as collaboration and information diffusion (Sutcliffe et al., 2012). What's more, observing the Ego Networks of different types of users can help to reveal insights about how they socialise. For example, various studies have compared the ENMs of specialised and generic users, the former being users who use a social media platform primarily for professional reasons and the latter being those that do not. It has been found that specialised users, in particular journalists, tend to have more fully formed Ego Networks on social media (Boldrini et al., 2018; Toprak et al., 2022b). This tells us that specialised users spend significantly more cognitive effort engaging with social platforms than generic users.

The ENM has also proved useful for a variety of network tasks, such as link prediction, where it has been used to extend preexisting prediction algorithms and outperform cutting-edge alternatives (Toprak et al., 2022a), and information diffusion, where it has been leveraged to gain novel insights about the propagation of trusted information (Arnaboldi et al., 2017). Indeed, given how ubiquitous this model is in human social networks, it has great potential for an even wider range of future problems. One clear example is stance detection, which aims to predict the attitude of a given text towards a given target entity.

Recent advances in stance detection have produced a predictive model named Cross-Target Text-Net (CT-TN). Many Stance Detection methods focus purely on the textual features in order to predict a stance. However, CT-TN employed the use of both textual and network (i.e. social) features and was able to outperform several other cutting-edge models. All these examples clearly demonstrate the potential impact of understanding how individuals interact within a network and, thus the potential of the ENM, for myriad network research tasks.

However, despite all the additional insights the ENMs can provide, it is not without its limitations. Notably, one of the core components of the ENM is the relationship strength between an Ego and their Alters. As the Alters are sorted into circles based on this metric, how it is measured is clearly fundamental to its implementation. But how to measure the strength of a relationship is not a problem with a clear or objective solution. In response to this, many researchers have traditionally used a much less subjective and much easier to obtain metric: contact frequency (Arnaboldi et al., 2013; Boldrini et al., 2018; Dunbar et al., 2015). While it has been proven that this is an effective proxy metric (Gilbert and Karahalios, 2009), it fails to take into account any of the qualitative aspects of a relationship. Take, for instance, an individual with a close friend who lives far away and an angry neighbour who is constantly causing trouble. Even though the individual may interact with these people at a similar frequency, the two relationships are clearly different in innumerable ways.

In Granovetter, 1973, a published definition of relationship strength was given as the equally weighted combination of 4 elements: the time spent maintaining it, its emotional intensity, its level of intimacy and the reciprocal services it generates. Clearly, the amount of time spent on a relationship will likely correspond to the number of interactions, which would explain why this proxy measurement has worked well historically. However, the other 3 elements of this definition, and all the additional information they could provide, are largely ignored when simply using contact frequency. So, there is clearly a very real gap between the information available within social networks themselves and the information retained by their graphical representations.

To address this informational shortcoming, this thesis proposes leveraging one of the fastest developing areas of computer science, Natural Language Processing (NLP), to add an additional dimension of qualitative detail to the ENM (for precise details, see Chapter 3). Specifically, this is attained by analysing the sentiments of each individual Ego-Alter relationship and using these sentiments to convert an ENM into a Signed Ego Network Model (SENM). This combination allows many of the useful properties of signed networks to be employed to improve our knowledge of the ENM and vice versa.

1.3 SIGNED NETWORKS

In a signed network, each edge has a polarity, either positive or negative. Within social contexts, positive links usually indicate friendship, trust and homogeneity (Maniu, Abdessalem, and Cautis, 2011), while negative links are associated with hatred and distrust. Such information can be easily used to improve the performance of many network tasks, such as community detection (Traag and Bruggeman, 2009), information diffusion (Ferrara and Yang, 2015) and opinion dynamics (Shi et al., 2016). Perhaps unsurprisingly, positive and negative relationships have been shown to play very different roles within a network, with negative links often providing significantly more information than positive ones (Leskovec, Huttenlocher, and Kleinberg, 2010a).

Given their relative usefulness, it is unfortunate that negative links are often less common in social networks, compared to their positive counterparts. Previous observations of networks with publicly available signed connections have found that negative connections tend to be much rarer than positive connections, making up around 15.0% to 22.6% of the total connections in a network (Leskovec, Huttenlocher, and Kleinberg, 2010b). The fact that link polarity is known to the users in these networks might amplify social pressure and effects such as social capital (Coleman, 1988), whereby relationships between individuals who have many relationships in common are more likely to be positive due to social pressure from the surrounding community. In unsigned networks, where negative connections are not explicitly visible, this social pressure should be lower. Hence, it would be reasonable to expect that networks without explicitly signed relationships would have higher proportions of negative relations than what is observed in explicitly signed ones. What's more, observing variations in negativity between different groups could reveal pertinent insights about the interactions and behaviours of those groups' members. This thesis explores such differences between different cultures and online communities in Chapter 6.

However, the extra information provided by signed networks goes beyond what can be gleaned merely from looking at signed edges themselves. When viewing a signed network as a whole, emergent properties can be observed. For example, signed networks are known to follow certain patterns, such as Balance Theory (Heider, 1946), which postulates that certain configurations of signed triads (i.e. groups of three users who are all interconnected by signed edges) should be more common than others when observed across an entire network. Specifically, triads with odd numbers of positive connections are considered plausible, or "balanced", while those with even numbers of positive connections are considered implausible, or "unbalanced". This is because connections are not independent but rather

influenced by the other connections in the surrounding network. An alternative version, known as weak Structural Balance Theory (Davis, 1967), also exists. This version is similar to the original Balance Theory but makes no assumptions about Triads with no or 1 positive connections. This latter version has been shown to better fit observations of online data. Balance Theory, which is an inherent expectation of signed social networks, has been used to predict the effectiveness of endorsements (Mowen and Brown, 1981) and for identifying subgroups within a network (Sun et al., 2020).

Unfortunately, due to a variety of moral and business reasons, the vast majority of social network platforms do not allow users to create explicitly negative links between themselves. Some explicitly signed datasets do exist, such as two of the most popular benchmark datasets for signed networks collected from Slashdot and Epinions (Tang et al., 2016). However, on top of being somewhat old for social media data, the information provided by these datasets is very limited; only providing user IDs and signs and not, for example, interaction frequencies or text associated with the interactions. Thus, there is a real need for up-to-date signed datasets. One solution that could address the lack of data could be to take an unsigned dataset and try to use it to synthesise the signs of the connections. Of course, this is far from trivial in practice and is a research challenge in and of itself.

Some methods have already been developed to predict the signs of unsigned networks. The majority of these have focused on using the structural elements of the surrounding network to determine the signs of relationships, for example, clustering coefficient (Javari and Jalili, 2014). Similarly, methods have been established for signing a novel network using classification algorithms trained on previous datasets with known signs (Tang et al., 2016; Yuan et al., 2019). All these techniques have viewed the problem of sign prediction from a top-down perspective, looking at the features of a network as a whole and inferring signs based on the structure of the links. However, by taking the inverse approach, and viewing the problem from the bottom-up, it is possible to consider the more qualitative aspects of connections that have largely gone unexplored.

Given that connections in most social media datasets are usually only binary (either two users are connected or not), one might assume that almost all of these connections would be positive. However, as previously discussed, this is not the case. Further, just as with real-world relationships, the online interactions of individuals usually contain implicit information about whether their connections are positive or negative (Maniu, Abdessalem, and Cautis, 2011). This means that the sign of a relationship in such online social networks can be thought of as an implicit piece of information that can be estimated from its corresponding interactions. Indeed, individual interactions are the basic blocks that constitute a relationship and looking at their indi-

vidual sentiments could reveal a lot about the relationship as a whole. What's more, sentiment analysis for individual texts is extremely well established (Liu, 2012) and it is possible to apply signs to these bottom-level interactions with a high degree of confidence. In addition to this, online social networks often provide a plethora of text-based communications that can easily be converted into signs via sentiment analysis.

However, ways of extending these singular labels to whole series of interactions, or relationships, are largely undeveloped in social network research. One previous study that has examined this problem (Hassan, Abu-Jbara, and Radev, 2012) employed a Support Vector Machine (SVM) to sign relationships between users in discussion forums based on 4 user features and 3 interaction features. The SVM was trained on an annotated dataset and is shown to achieve an accuracy above 0.835 on the test set. Unfortunately, this approach is not directly and reliably replicable for many social media platforms, which are typically characterised by shorter and less structured interactions than forum discussions. In addition, the ground truth dataset used for the training phase is not publicly available, making it extremely difficult to reproduce this method.

1.4 THESIS CONTRIBUTIONS AND STRUCTURE

This chapter has highlighted the need for novel signed network data as well as the high availability of text-based interactions and their suitability for sentiment analysis. It has also provided an overview of the ENM and its supporting literature, while also emphasising a notable limitation of this model: the way in which the cornerstone concept of tie strength has traditionally been measured. The main aims of this thesis are:

1. To establish a methodology of converting unsigned networks into signed networks using text-based interactions between users.
2. To apply this novel signing method to the ENM, thus introducing the SENM.
3. To discover new insights about the ENM resulting from the addition of signed connections.
4. To explore potential applications of the SENM.
5. To observe differences in the SENM between different cultural groups and online communities.

The work conducted throughout the course of this thesis has also resulted in 5 publications: 2 journal papers and 3 conference papers. These publications are listed in Appendix A.1.

The above contributions have been obtained by collecting and analysing a significant number of datasets. Specifically, Chapter 2 lays the groundwork for this work by providing a comprehensive list of the numerous datasets used throughout this thesis. As these datasets are used repeatedly in various combinations throughout the subsequent chapters, they have been grouped together at the beginning to minimise repeated information or overly circuitous references.

Following the collection of datasets, Chapter 3 proposes a solution to the lack of signed social network data by presenting a novel approach to signing relationships based solely on text-based interactions. Results obtained using this approach are also tested using a variety of different sentiment analysis models, before being validated using known structural expectations of signed networks.

Using this method of signing relationships, Chapter 4 proposes an extension to the ENM, the SENM, that includes the signs of Ego-Alter relationships and, thus, adds an additional layer of important qualitative information to this model. The properties and distributions of signs within this new model are then explored, as well as the effects of negative relationships on cognitive effort.

In Chapter 5, both the ENM and SENM are applied to the task of stance detection. This chapter demonstrates a concrete use case for the additional information these models provide, supporting their potential for impacting further networking tasks.

In the penultimate chapter, Chapter 6, variances in the SENM are compared between different cultural groups and online communities. Follow-up investigations of the words and topics used by different groups, relative to their levels of negativity, are also conducted.

Finally, a summary of all major conclusions as well as potential avenues of future research are discussed in Chapter 7.

2

DATASETS

A total of 35 datasets are used in this thesis, representing a mixture of 11 pre-existing and 24 novel sets, the latter being collected specifically for SENM analysis. These datasets are used in repeating in varying combinations throughout the subsequent chapters. Thus, for clarity, all information pertaining to the datasets is explained here and all future chapters refer back to this chapter, specifying which datasets are relevant.

First, the data source and its API are introduced, followed by descriptions of the datasets used in the thesis, then the preprocessing steps required to correctly prepare the data and, finally, analyses of the unsigned Ego Networks, to ensure they are in line with expectations based on previous findings.

2.1 DATA SOURCE

All data used in this thesis was collected from the X social media platform (formerly known as Twitter). X has long been a reliable source for Ego Network research due to its vast and active userbase (with between 100 and 250 million daily users globally during the period in which data was collected for this thesis (Statistica, 2022)).

What's more, although users can change the level of privacy of their account, the vast majority of posts, known as Tweets, are public. These Tweets are comprised of up to 280 characters and can also be used to tag other users. Indeed, Tweets contain information about their author and any other users associated with them, which makes it very easy to identify communications between users and to assign an Ego and an Alter (or Alters) to each one. Such interaction Tweets can be sorted into 3 categories: Replies, when a user specifically published a Tweet in response to one made by another user, Mentions, when a user tags another user using their username, and Retweets, when a user shares a Tweet from another user. This latter category is slightly different from the former two as it allows users to "interact" by merely clicking a button. When making a Retweet, users can also add some text of their own (known as a Quote Retweet). In order to better reflect the cognitive effort required for maintaining a relationship, only Quote Retweets are considered in this work. Similarly, only the interactions created by the Ego are used when computing the frequency of interaction with their Alters. This is done to better reflect the cognitive and time constraints of the Ego.

In addition to the suitability of the data, at the time of data collection, X also provided a very accessible API that was free for academic users[1]. The API allowed data to be collected in a variety of ways, including 2 endpoints that were particularly important for the collection of the data used in this thesis: X Search and User Timelines.

X Search

By using X Search, it was possible to search all of X for a specific query and to receive a stream of related Tweets in reverse chronological order (i.e. newest first). This made it possible to easily discover users who had recently engaged on specific topics and, when combined with a geographically distinct query, to limit the search to users connected to specific cultures or regions. Further details on the collection methodology of each dataset are available in their corresponding subsections later in this chapter.

User Timelines

After identifying a list of users to be collected, it was possible to get the User Timeline of each one based on their unique ID or screen name. A Timeline is a collection of all of the public Tweets a user has posted throughout the entirety of their time on X. However, at the beginning of the 6-year time span during which the novel datasets were collected (2018 to 2023), there was a limit on the total number of Tweets that could be obtained via User Timelines. As such, the exact amount of data obtained for each user varies slightly, depending on which version of the API was available at the time of collection. Specifically, the user timelines gathered using version 1 (v1) of the API were limited to the most recent 3,200 Tweets, whereas the length of timelines gathered using version (v2) had no upper limit. Although a user may have more than 3,200 tweets, previous findings (Arnaboldi et al., 2017, 2015; Dunbar et al., 2015) as well as the results of the unsigned Ego Network analysis conducted on these datasets, see Section 2.5, suggest that this number is more than sufficient to develop a complete Ego Network. With the exception of P-Stance (see Section 2.2), all of the datasets used in this thesis, including the pre-existing ones, are comprised of a collection of User Timelines, with each Timeline corresponding to a single Ego.

2.2 PRE-EXISTING DATASETS

The 11 datasets taken from prior work include 7 on journalists from various regions, 1 set containing science fiction writers, 2 generic

1 Unfortunately, as of July 2023, the Academic API is no longer available.

sets of users and 1 on users involved in the US 2020 presidential election. These represent a variety of both generic and specialised users. Recalling from Section 1.2 that these two types of users have previously displayed notable differences in behaviour, especially with regards to the ENM. It is, therefore, important to make a specific distinction between them. All of these were collected using v1 of the X API.

Journalists

Most of the journalist datasets are originally from a study that investigated the Ego Networks of journalists from all across the globe (Toprak et al., 2021). These journalists use the X platform primarily for professional reasons. They were collected between January and May 2018 using manually curated lists of users, which were also validated by the previous researchers (Boldrini et al., 2018). These include journalists from Australia, Brazil, Italy, the Netherlands, the United Kingdom and the United States of America. Of course, half of these datasets contain Tweets in non-English languages, adding another layer of complexity to the problem.

In addition to the above, another set of journalists was taken from a different study (Ollivier et al., 2022). This dataset contains journalists from the New York Times (NYT) and the list of users was created by the New York Times itself. This dataset was originally collected in February 2018.

Science Writers

The users included in the Science Writers dataset come from a list created by a science writer at Scientific American, Jennifer Frazer. Similarly to the journalists, these are users who use X professionally, albeit to a potentially different extent. These timelines were gathered in June 2018, as part of the same study as the NYT journalists (Ollivier et al., 2022).

Generic Users

Two types of generic users were gathered from previous datasets. The first of these is users who Tweeted in English using the hashtag #MondayMotivation on 16th January 2020. This hashtag is often used to share motivational images or stories and is, therefore, used by a wide variety of users. The second is a random sample of users who posted English Tweets from the United Kingdom on the 11th of February 2020. These datasets were collected between January and

February 2020 as part of the aforementioned NYT and Science Writers work (Ollivier et al., 2022).

P-Stance

This dataset is notably distinct in nature from the others because is not a collection of User Timelines. Rather, P-Stance contains 21,574 English-language tweets collected during the 2020 U.S. presidential election (Li et al., 2021). These tweets were specifically collected to be used for SD and, because of this starkly different nature (i.e. because it does not contain User Timelines), this dataset was only used for SD in this paper and not for any sort of Ego Network analysis. Each of the P-Stance Tweets is associated with one of 3 targets, Joe Biden, Donald Trump or Bernie Sanders, and a corresponding stance label, either "AGAINST" or "FAVOR". Hashtags, such as "#BidenForPresident" and "#NeverBernie" were used to both search for the tweets and to determine their target and stance. While the original dataset only included the text and stance of each tweet, the authors of the P-Stance dataset provided 9,307 tweet IDs upon request, allowing further data to be collected for each tweet, including information about the authoring users. In addition to providing the information required for computing the users' Ego Networks, this also made it possible to obtain, for each user, the remaining features required by the CT-TN model: likes, followers and friends.

2.3 NOVEL DATASETS

The 24 novel datasets consist of a British MPs dataset, a baseline dataset, 4 sets of generic users grouped by geographical region, 3 Reality TV datasets, 4 of users engaged in politics, 4 engaged with football, 4 with the weather and 3 further sets of country-specific generic users. These datasets were specifically selected to better explore the interplay between negativity and engagement. So, while all but the British MPs represent generic users, each set of datasets is concentrated around specific topics with varying levels of polarisation and engagement. These datasets are presented in the section in the order they were collected.

British Members of Parliament

A collection of 584 British Members of Parliament (MPs) was collected using a publicly available list of official MP X accounts, provided by UKinbound, 2020. This list includes 594 MPs, however, given that the data collection happened 2 years after the publication of the list, the users were manually checked to see if they were still active, both

on X and in parliament. Thus, 10 users had to be removed. The User Timelines of the remaining individuals were collected in March 2022, using v1 of the X API[2].

Baseline

The Baseline dataset was the first to be collected as part of this work. Its purpose, as its name suggests, was to gauge a baseline level of negative relationships on X. In addition, this dataset was used to validate the signing methodology proposed by this thesis. In order to perform this validation, the Baseline dataset needed to have a reasonable amount of interconnectivity between users, i.e. the Egos had to be connected to one another (see further details in Section 3.4). Therefore, a snowball sampling methodology was employed.

This involved first identifying a small subset of initial "seed" Egos that was small enough to be manually vetted, then collecting those users' Alters, followed by the Alters' Alters. These seeds were selected from an extremely large, pre-existing dataset of Ego Networks (itself a snowball dataset, collected in November 2012, that used Barack Obama as the initial seed) (Arnaboldi et al., 2013). First, the most connected user was chosen, then, a further 30 users were pseudorandomly selected, with the prerequisite of being connected to at least one of the previously chosen users; this resulted in 31 seed users. This approach allowed for large numbers of users to be collected, far beyond what was possible to check manually, while still being reasonably sure that the majority of users were well-connected, active users who regularly engage with X. As the original dataset was collected over a decade ago, the User Timelines of the selected seeds had to be recollected to ensure their Alters were up to date. Thus, the Baseline dataset was collected between April and May 2022 using the v1 API[3].

After collection, all the included users' locations were investigated to get a measure of how diverse a sample they represented. Users on X may choose to explicitly add a location to their profile, however, this is done using a text field, in which they may write anything they like. As such, many user profiles have locations that are empty, undecipherable or even fictitious (e.g. "Narnia"). Despite this, it is possible to identify real-world locations for users by passing their self-defined locations into the Google Maps API and checking which ones actually exist and where they are. In this manner, after performing the preprocessing steps described in Section 2.4, locations were able to be obtained for 17.5% of the users. These locations span 68 countries and the numbers and percentages of users belonging to them can be seen in Table 1 (excluding countries with fewer than 3 identified users). The population is notably skewed towards the US and UK, and, while

2 Tweet IDs of this dataset are available at `https://zenodo.org/record/6420845`.

3 Tweet IDs of this dataset are available at `https://zenodo.org/record/7717006`.

Table 1: The numbers and percentages of users in the Baseline dataset, for whom it was possible to confirm a real-world location, grouped by country (excluding countries with fewer than 3 identified users).

Country	Number of Users	Percentage
United States	899	70.45
United Kingdom	129	10.11
Nigeria	61	4.78
Canada	60	4.70
Australia	17	1.33
Spain	15	1.18
Brazil	13	1.02
South Africa	10	0.78
Ireland	8	0.63
India	8	0.63
Netherlands	8	0.63
Mexico	6	0.47
Sweden	6	0.47
Italy	6	0.47
Germany	5	0.39
France	5	0.39
Belgium	4	0.31
Colombia	4	0.31
Jamaica	4	0.31
Jordan	3	0.23
Japan	3	0.23
Poland	3	0.23

this is also reflected in the X userbase itself (Statistica, 2024), it could suggest a bias towards anglophone countries in the Baseline dataset. This is further highlighted by the fact that Japan, which has the second highest number of active X users (Statistica, 2024), only represents 0.23% of the locations identified in Baseline.

Regional Datasets

Following the collection of the baseline dataset, similar datasets were collected for specific regions, with the purpose of comparing observations between different cultures. Thus, the same snowball sampling methodology was employed using initial seeds from a variety of dif-

ferent regions. As the number of active X users varies enormously from country to country, it is not possible to collect meaningfully large datasets for every region of the globe. Therefore, regions were prioritised if they had large active userbases, informed by the numbers of users in the Baseline dataset (see Table 1). In total, 4 culturally distinct parts of the world for which data was expected to be relatively available were identified. These were the Mediterranean (Spain, France, Italy, Greece), South America (Brazil, Colombia, Venezuela), Northern Europe (Germany, Netherlands, Sweden) and West Africa (Nigeria, Senegal, Ghana). These regions also had the added benefit of corresponding fairly well with the availability of journalist data, described in Section 2.2. Indeed, the only region for which there was no journalist data available was West Africa.

Given the aforementioned problems when identifying users' locations, particular attention was paid to the selection of the regional datasets' initial seed users. For each region, 3 seed users were identified per country. These users were selected by querying the X Search endpoint for the name of each country in its most widely spoken official language (for example, "España", "Deutschland" or "le Sénégal") and then randomly sampling from the users who had posted within the previous 24 hours. Before being included, each seed was manually checked to ensure that they were generic users (i.e. users that don't use X for professional reasons) from the desired country. Of course, it was not possible to manually check every user that was collected, so this was only done for the initial seeds. The data were collected between June and July 2022 using the v1 API[4].

Reality TV

Next, a further set of datasets were collected with the express purpose of finding an approximate midpoint between journalists and generic users, in terms of level of engagement. Communities based around reality TV were selected as such shows tend to have large, highly committed audiences, while still being mostly generic in terms of content. Additionally, many countries have shows that are either unique to them or specifically localised, making it very easy to match these communities to specific countries. Unfortunately, no journalist dataset corresponding to the region of West Africa was available, so data were only collected for the other 3 regions. As reality TV shows tend to be specific to individual countries, rather than larger geographic regions, only the most populous (in terms of X users) country from each region was used. As can be seen in Table 1, the most populous country from each of the selected regions were Italy, Brazil and the Netherlands (and would have also been Nigeria if journalists were available).

4 Tweet IDs of these datasets are available at `https://zenodo.org/record/7717047`.

Again, a snowball sampling was used. This time, hashtags related to specific reality TV shows were used to search and identify users engaging in these topics. For Italy and Brazil, the local versions of "The X Factor" (#XF2022 and #XFactorBR) and "Big Brother" (#GFVIP and #BBB22) were used. For the Netherlands, these shows have the same titles as their British and American counterparts. So, to avoid collecting users from other countries, similar, non-linguistically ambiguous TV shows were chosen: "Holland's Got Talent" (#HollandsGotTalent) and "Ik Vertrek" (#IkVertrek). For each hashtag, 3 manually vetted seed users were collected. While there is no guarantee that the non-seed users collected in this manner won't spend the majority of their time tweeting about other topics, a look at the most common topics discussed by these users reveals that this does not seem to be the case. Indeed, most of their communications are discussing topics related to reality TV (see Subsection 6.2). These data were collected in January 2023 using v1 of the API[5].

Politics

Next, further datasets of various types of users were collected by country in order to be able to make more fine-grained observations between both culture and type of community. In addition to Italy, Brazil and the Netherlands, these datasets also include Nigeria. The first of these community-based collections gathered users who were engaging in political discussions. The seeds were users who commented on posts made by political parties in each of the target countries. In order to get a broader impression of the general political discussions of each country, rather than that of any single political party, a list of seed users was generated for multiple parties for each country, and the final seeds used were selected randomly from these lists and then manually vetted for suitability, with a minimum of 5 users from each party's list. For Italy, the chosen parties were Fratelli d'Italia (@FratellidItalia), Lega Salvini (@LegaSalvini) and Partito Democratico (@pdnetwork), for Brazil they were Partido Liberal (@PartidoLiberal), Movimento Democrático Brasileiro (@MDB_Nacional) and Partido dos Trabalhadores (@ptbrasil), for the Netherlands Volkspartij voor Vrijheid en Democratie (@VVD), Democraten 66 (@D66), Christen-Democratisch Appèl (@cdavandaag) and for Nigeria they were All Progressives Congress (@OfficialAPCNg), Peoples Democratic Party (@OfficialPDPNig), Labour Party (@OfficialPDPNig) and New Nigeria Peoples Party (@OfficialNNPPng). These data were collected between April and May 2023 using the v2 API[6].

5 Tweet IDs of these datasets are available at `https://zenodo.org/record/7716860`.
6 Tweet IDs of these datasets are available at `https://zenodo.org/records/10605838`.

Football

The second collection of community-based users was centred around the topic of football. The seeds were users who commented on posts made by popular local football teams: Juventus (@juventusfc) for Italy, Regatas do Flamengo (@Flamengo) for Brazil, Ajax (@AFCAjax) for the Netherlands and Enyimba (@EnyimbaFC) and Plateau United (@plateau_united) for Nigeria. At least 3 seeds were used for each country, again, these were all manually vetted before being used. Nigeria was the only country whose most popular football teams were from foreign countries, such as Manchester United, Chelsea and Barcelona. This was also why a second team was included, as it was not possible to get enough seed users using a single Nigerian football team. These data were collected between April and May 2023, using the v2 API[7].

Weather

The next set of users is somewhat more generic and based around the topic of weather. At least 3 seeds were used per country, randomly selected from users who commented on the posts of local weather forecasting X accounts. These accounts were Meteo Italia (@meteo_italia7) and meteo.it (@wwwmeteoit) for Italy, MetSul Meterologia (@metsul) for Brazil and Weer & Radar Nederland (@weerenradar_nl) for the Netherlands. For Nigeria, it was not possible to find a weather-related account that generated more than a few comments from other users, so the seed users were instead collected using the X Search endpoint to search for "weather nigeria". As usual, all seed users were manually observed to ensure their suitability. These data were collected between April and May 2023, using the v2 API[8].

Countries

Finally, generic datasets were collected for each of the 4 target countries. These were collected between May and June 2023 using the v2 API. Unfortunately, the Academic track of the API was closed by X before all the datasets could be collected, meaning that the Nigerian dataset is missing[9].

Once more the snowball sampling methodology was used. Initial users were collected by querying the X Search endpoint for the names of the target countries in their official languages ("l'italia", "el brasil" and "het nederlands"). From the results of these searches, 3 interconnected users were randomly selected and manually vetted to ensure

7 Tweet IDs of these datasets are available at `https://zenodo.org/records/10605838`.
8 Tweet IDs of these datasets are available at `https://zenodo.org/records/10605838`.
9 Tweet IDs of these datasets are available at `https://zenodo.org/records/10605838`.

that they were generic users from the desired countries. These users were used as the seeds for the collection.

Because it is not possible to manually check the country of every user in the datasets and, as the snowball sampling method allows for the possibility of users to be collected from outside the target country, some tests were performed after the collection of these datasets. This involved checking the self-declared location and the main language of each user (provided by X). For the Brazilian and Dutch datasets, the vast majority of users' locations and languages were as expected. However, the Italian dataset showed a significant proportion (around a third) of users from the UK and USA, as well as smaller groups from other countries, such as France. To combat this, users were removed from this dataset if their location contained "UK", "London", "USA", "DC" or "CA" or if their main language was not Italian. Of course, this would remove any Italian users who prefer to use X in English but it was deemed to be the best way to ensure the removal of non-Italian users. Unfortunately, this resulted in a much smaller dataset than the other two Generic Users, however, it is still one of the largest datasets used in this thesis (see Table 2 in Section 2.4). Indeed, the 3 generic datasets that were able to be collected were specifically made much larger than the other country-specific datasets. This was done in order to investigate whether any observable differences between communities that had been specifically targeted during collection could also be observed between the naturally occurring communities within a larger collection (see Section 6.2 and Section 6.3).

2.4 PREPROCESSING

While X is known to be an excellent source of social media data, it is not without some problems. Indeed, there is a large number of users that are not individual humans, are inactive or have irregular activity, such as bots or accounts controlled by businesses. In the former case, groups of humans working together will obviously have very different cognitive constraints compared to individuals, and bots will not have any. In the latter, users who are inactive or have erratic activity are unlikely to be engaged enough to have fully developed Ego Networks on the platform. Therefore, it is important to remove any such accounts before analysing the users' Ego Networks.

After the removal of non-human and irregular users, the remaining data correspond to the Egos' full networks (i.e. every human user the Ego has interacted with throughout their time on X). While, after also removing the inactive users, they correspond to the active part of the Ego Networks (i.e. the part of the Ego's network comprised of only the relationships with which they interact at least once a year, as per the ENM). Descriptive statistics of all the datasets, except P-Stance (due to its significantly different nature), after the preprocessing steps,

are displayed in Table 2, grouped by type and ordered by date of collection.

Non-Human Users

First, non-individual human users were removed using a Support Vector Machine (SVM) (Cortes and Vapnik, 1995), trained on a sample of 500 X users. This approach has previously been established in ENM research (Arnaboldi et al., 2013). The users in the training set have been manually labelled as either "people" or "other" and the data contain 100 timeline-related features and 15 profile-related features for each of them. In its initial publication, the SVM achieved an accuracy of 81.3% using k-fold cross-validation (with k=5), with a false positive rate below 10%[10]. After training the model, it is then a trivial matter to apply it to new accounts and remove any that it predicts as "other".

This step only had to be performed for certain datasets. Specifically, it was not done for the Journalist, Science Writers or British MPs, as the users in these datasets were gathered from manually curated lists of known individuals (refer to the details of each dataset in Sections 2.2 and 2.3 for more details). This step was also not performed for the P-Stance dataset as it was not used for Ego Network analysis.

Inactive and Irregular Users

Next, users who did not display significant and regular levels of activity in the months leading up to data collection were filtered out. This was defined as Egos who had timelines that spanned fewer than 6 months, that had less than 2,000 Tweets total (including non-interaction Tweets) or who tweeted less than once every 3 days for the majority of the months they were included in the dataset. Additionally, relationships that had fewer than 1 interaction per year were excluded. These filtration parameters are in line with those of previous work on Ego Networks (Arnaboldi et al., 2015; Toprak et al., 2021) and stem from psychological and anthropological research (Hill and Dunbar, 2003).

Unlike the removal of non-humans, this preprocessing step had to be performed on all the datasets, even those made from lists of verified individuals. This is because some users may have become inactive or changed their behaviour between the time when the lists were published and when the data was collected.

10 Meaning that less than 10% of the non-humans were erroneously labelled as humans.

Table 2: Number of Egos, Alters and interactions in each dataset (after all preprocessing steps)

Dataset	Egos	Alters	Interactions
American Journalists	1,037	68,792	1,639,623
Australian Journalists	520	26,561	937,764
British Journalists	281	24,614	434,477
Italian Journalists	203	14,192	489,008
Brazilian Journalists	154	11,580	278,631
Dutch Journalists	1,316	54,377	2,702,275
NYT Journalists	558	23,327	561,563
Science Writers	241	18,531	381,340
Monday Motivation	1,461	78,906	894,648
UK Users	921	84,993	1,474,882
British MPs	584	157,053	1,277,010
Baseline	4,049	346,168	8,593,290
Mediterranean	878	56,326	2,191,666
South America	217	13,458	441,158
Northern Europe	552	42,090	1,273,881
West Africa	396	27,705	884,321
Italian Reality TV	160	11,191	291,213
Brazilian Reality TV	154	9,044	234,734
Dutch Reality TV	230	14,095	441,694
Italian Politics	2,004	50,579	1,744,040
Brazilian Politics	482	17,112	354,664
Dutch Politics	1,256	47,399	1,215,376
Nigerian Politics	866	21,291	526,768
Italian Football	1,320	35,983	1,132,520
Brazilian Football	1,024	42,759	937,592
Dutch Football	1,910	66,496	1,248,816
Nigerian Football	159	8,494	113,648
Italian Weather	518	22,927	337,344
Brazilian Weather	598	22,650	340,424
Dutch Weather	255	10,427	115,416
Nigerian Weather	363	11,616	181,024
Italian Generic	2,740	150,682	2,133,608
Brazilian Generic	8,223	404,396	6,561,320
Dutch Generic	9,278	404,484	6,905,496

2.5 EGO NETWORK ANALYSIS

After collecting all the data, the Ego Networks of each user were analysed to make sure that they were in line with expectations from previous research. As such, these analyses are more of a confirmation that the data is appropriate for work with ENMs, rather than a contribution of anything novel. Because the pre-existing dataset only contained a collection of User Timelines, rather than the previously computed Ego Networks, this confirmation step was done for all the datasets, not just the novel ones. This provided an additional check on the suitability of the code used for the computations of the ENMs.

Computing Ego Networks

After collecting and preprocessing all of the data, the Ego Network of each user could be computed by clustering their Alters based on their Ego's frequency of contact with them. Specifically, this is done by first matching all of the interactions in an Ego's User Timeline to each of their corresponding Alters and calculating their interaction frequencies. Then, a clustering algorithm can be used to group the Alters into their appropriate circles. Previously, this has been done with a variety of different clustering algorithms; including k-means (MacQueen, 1967; Arnaboldi et al., 2013), DBSCAN (Ester et al., 1996; Arnaboldi et al., 2012), MeanShift (Fukunaga and Hostetler, 1975; Boldrini et al., 2018) and the Jenks breaks (Jenks, 1967; Paraskevopoulos et al., 2021). However, the method used in this thesis employed a combination of MeanShift and the Jenks breaks. The Jenks breaks algorithm seeks to minimise each class' average deviation from its mean, making it especially appropriate for dealing with 1-dimensional data, such as interaction frequencies. However, it is not able to determine an appropriate number of clusters on its own. Hence, MeanShift, which is able to determine an optimum number of clustering groups but works best on data with at least 2 dimensions, was used first to identify the optimum number of circles for each Ego and this number was then used to run Jenks breaks. The mean optimum number of circles and Ego Network sizes of each dataset can be seen in Table 3.

Analysing Ego Networks

Following the computation of the Ego Networks, the mean sizes of each circle, displayed in Table 4, were observed and compared to expectations from previous research. It is important to note that the sizes of Ego Networks and their circles tend to vary slightly from Ego to Ego, due to various social differences between individuals. Because of these common variations, and in order to standardise the results of any analysis performed on the circles, it is standard practice to focus

Table 3: Mean optimum number of circles and Ego Network sizes in the active networks of each dataset

Dataset	Optimum # Circles[a]	Ego Network Size[a]
American Journalists	5.30 [5.22, 5.38]	138.27 [134.26, 142.29]
Australian Journalists	5.17 [5.06, 5.28]	145.11 [139.55, 150.66]
British Journalists	5.42 [5.27, 5.56]	147.77 [140.47, 155.08]
Italian Journalists	5.72 [5.20, 5.93]	120.10 [110.49, 129.70]
Brazilian Journalists	5.46 [5.27, 5.65]	116.48 [103.95, 129.01]
Dutch Journalists	5.45 [5.42, 5.51]	122.69 [118.42, 126.96]
NYT Journalists	5.53 [5.40, 5.66]	155.67 [148.61, 162.74]
Science Writers	5.44 [5.29, 5.60]	148.45 [148.45, 156.77]
Monday Motivation	5.07 [5.00, 5.15]	107.42 [103.83, 111.00]
UK Users	5.23 [5.14, 5.32]	120.98 [115.75, 126.22]
British MPs	6.00 [5.87, 6.12]	174.68 [168.13, 181.22]
Baseline	4.81 [4.78, 4.84]	99.05 [96.49, 101.60]
Mediterranean	5.11 [5.03, 5.18]	109.68 [103.78, 115.58]
South America	4.85 [4.73, 4.96]	101.94 [92.89, 110.99]
Northern Europe	5.10 [5.01, 5.19]	119.60 [112.68, 126.51]
West Africa	5.00 [4.90, 5.10]	102.32 [94.94, 109.69]
Italian Reality TV	5.48 [5.25, 5.71]	103.65 [88.05, 119.25]
Brazilian Reality TV	5.42 [5.22, 5.62]	96.63 [85.11, 108.14]
Dutch Reality TV	5.20 [5.01, 5.39]	98.63 [86.41, 110.85]
Italian Politics	5.12 [5.06, 5.18]	104.00 [99.49, 108.51]
Brazilian Politics	4.91 [4.80, 5.03]	100.88 [92.76, 109.01]
Dutch Politics	5.24 [5.16, 5.32]	111.31 [104.63, 118.00]
Nigerian Politics	4.67 [4.59, 4.75]	87.55 [82.27, 92.83]
Italian Football	5.08 [5.01, 5.15]	107.44 [101.74, 113.13]
Brazilian Football	5.27 [5.20, 5.35]	107.01 [101.74, 113.13]
Dutch Football	4.66 [4.60, 4.72]	88.65 [83.82, 93.48]
Nigerian Football	4.80 [4.59, 5.01]	110.76 [96.58, 124.93]
Italian Weather	4.67 [4.55, 4.78]	102.86 [91.11, 114.62]
Brazilian Weather	4.54 [4.45, 4.78]	90.17 [81.54, 98.80]
Dutch Weather	4.28 [4.13, 4.43]	88.06 [73.50, 102.63]
Nigerian Weather	4.36 [4.24, 4.49]	81.60 [70.89, 92.31]
Italian Generic	4.88 [4.83, 4.94]	101.48 [97.36, 105.60]
Brazilian Generic	4.96 [4.93, 4.99]	101.86 [99.60, 104.11]
Dutch Generic	4.79 [4.76, 4.82]	102.64 [100.12, 105.16]

[a]95% confidence intervals shown in square brackets.

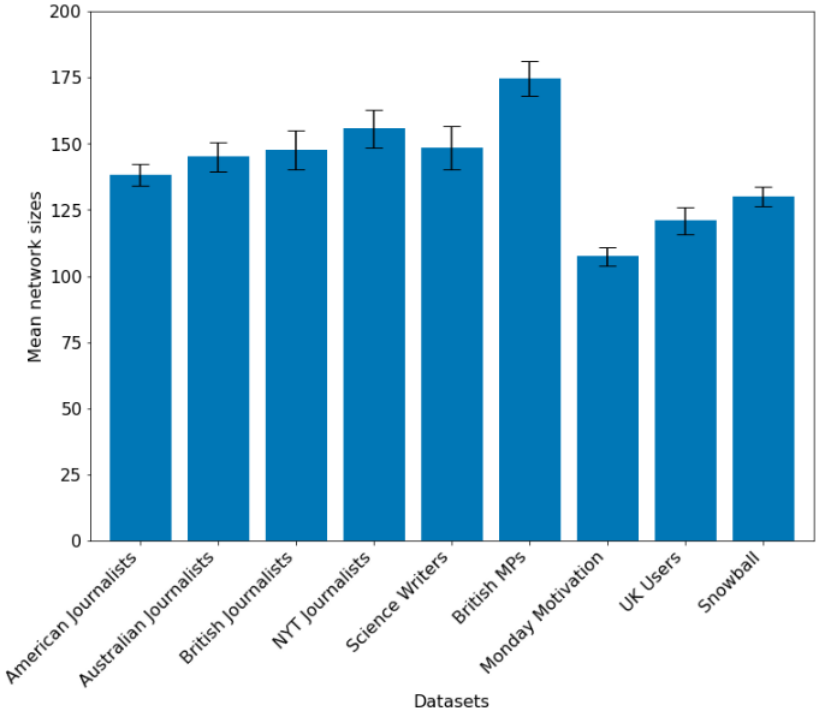

Figure 3: Mean active Ego Network sizes of users with 5 circles in each dataset (95% confidence intervals)

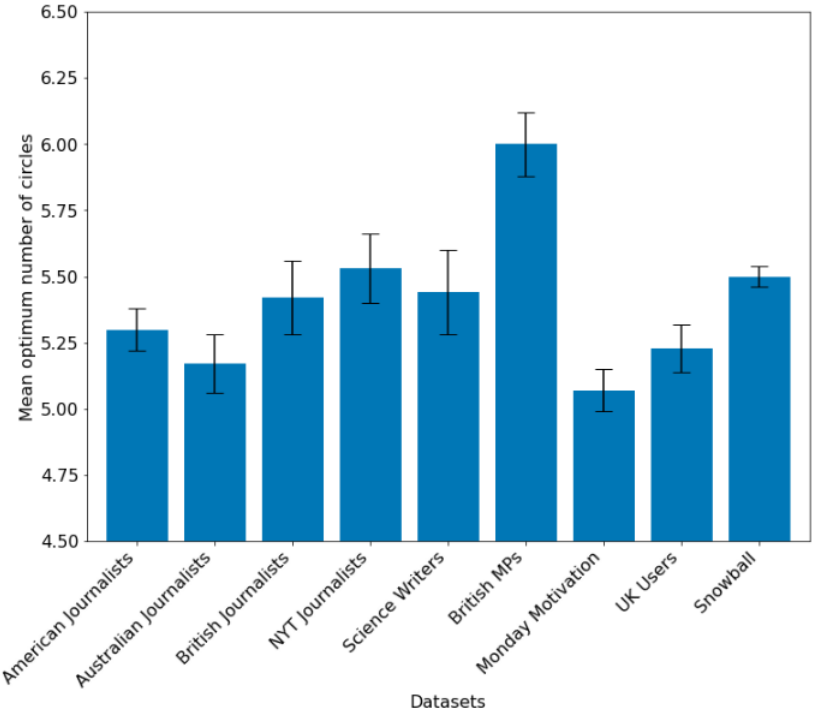

Figure 4: Mean number of circles for each dataset (95% confidence intervals)

on Egos with a common number of circles (Boldrini et al., 2018; Toprak et al., 2021). Usually, the chosen number of circles is 5 as it is the most common number for OSN data (Arnaboldi et al., 2017; Toprak et al., 2022a) and, as can be seen, 5 is the closest whole number for the vast majority of the datasets. The only exceptions are Italian Journalists, NYT Journalists and British MPs, for which the closest whole number is 6, and Dutch and Nigerian Weather, for which it is 4. This suggests that 5 is also the most appropriate number for the datasets included in this thesis and users with 5 circles will, therefore, be the main focus of ENM analyses herein.

Recalling from Chapter 1 that the expected circle sizes are 1-2, 5, 15, 45-50 and 150, with values from online data expected to be slightly lower for the outer circles, the values observed for all datasets in this thesis appear extremely close to expectations. Indeed, all of the means for circle 1 are within the expected range, all of the second circles are between 4 and 6 and the third circles are all between 10 and 20. While a handful of the outer circles are towards the lower limit of what could reasonably be expected, such a drop-off has previously been observed in online data and can be attributed to users having Ego Networks that are not completely formed online (Arnaboldi et al., 2017; Toprak et al., 2021). What's more, the scaling ratios between each circle are displayed in Table 5 and the expected factor of around 3 is clearly visible for all datasets.

Thus, all of the datasets gathered for use in this thesis, both preexisting and new, display Ego Networks that are very much in line with expectations based on previous literature. This includes the sizes of the active networks and the number of circles in the Ego Networks as well as the sizes and scaling ratios of these circles. The datasets, therefore, appear to be suitable for ENM analysis.

Table 4: Mean circle sizes of Egos with an optimum circle number of 5

Dataset	Circle 1	Circle 2	Circle 3	Circle 4	Circle 5
American Journalists	1.61	5.33	15.01	41.78	127.28
Australian Journalists	1.41	4.76	13.61	40.22	134.71
British Journalists	1.83	6.27	16.87	48.07	142.52
Italian Journalists	1.12	3.57	10.59	33.14	120.10
Brazilian Journalists	1.78	5.90	15.66	41.62	116.48
Dutch Journalists	1.66	5.51	15.54	43.08	122.69
NYT Journalists	1.65	5.43	14.76	40.16	114.68
Science Writers	1.70	5.81	16.40	44.29	124.86
Monday Motivation	1.72	5.26	13.22	33.58	103.71
UK Users	1.84	5.96	15.72	39.32	114.66
British MPs	1.98	6.67	18.09	49.00	146.79
Baseline	1.78	6.16	16.86	44.19	125.91
Mediterranean	1.70	5.60	14.67	38.83	120.41
South America	1.80	5.76	15.71	39.92	118.29
Northern Europe	1.80	5.88	17.12	45.34	131.12
West Africa	1.65	5.60	15.64	39.71	118.81
Italian Reality TV	1.63	5.15	13.85	35.60	103.65
Brazilian Reality TV	1.58	4.65	12.08	31.29	96.63
Dutch Reality TV	1.61	5.29	14.29	37.16	98.63
Italian Politics	1.70	5.38	13.40	33.38	101.22
Brazilian Politics	1.70	5.07	13.31	33.11	97.55
Dutch Politics	1.81	5.77	14.72	36.85	108.59
Nigerian Politics	1.52	4.56	11.47	27.70	83.22
Italian Football	1.72	5.52	14.12	35.32	104.91
Brazilian Football	1.63	5.24	13.43	34.67	106.67
Dutch Football	1.56	4.84	12.27	29.51	86.15
Nigerian Football	1.54	5.22	12.49	32.07	101.24
Italian Weather	1.63	4.97	12.81	32.28	94.38
Brazilian Weather	1.46	4.49	11.20	27.73	80.27
Dutch Weather	1.34	4.29	10.41	25.80	75.96
Nigerian Weather	1.35	4.18	10.38	24.00	71.25
Italian Generic	1.55	4.91	12.77	32.43	96.29
Brazilian Generic	1.63	5.00	12.98	33.31	98.29
Dutch Generic	1.62	5.09	13.25	33.76	97.11

Table 5: Scaling ratios between circle sizes

Dataset	Circle 1-2	Circle 2-3	Circle 3-4	Circle 4-5	Mean
American Journalists	3.30	2.82	2.78	3.05	2.99
Australian Journalists	3.37	2.86	2.96	3.35	3.13
British Journalists	3.43	2.69	2.85	2.96	2.98
Italian Journalists	3.19	2.97	3.13	3.62	3.23
Brazilian Journalists	3.31	2.65	2.66	2.80	2.86
Dutch Journalists	3.32	2.82	2.77	2.85	2.94
NYT Journalists	3.30	2.72	2.72	2.86	2.90
Science Writers	3.41	2.82	2.70	2.82	2.94
Monday Motivation	3.07	2.51	2.54	3.09	2.80
UK Users	3.24	2.64	2.50	2.92	2.82
British MPs	3.37	2.71	2.71	3.00	2.95
Baseline	3.46	2.74	2.62	2.85	2.92
Mediterranean	3.29	2.62	2.65	3.10	2.92
South America	3.21	2.73	2.54	2.96	2.86
Northern Europe	3.26	2.91	2.65	2.89	2.93
West Africa	3.38	2.80	2.54	2.99	2.93
Italian Reality TV	3.17	2.69	2.57	2.91	2.84
Brazilian Reality TV	2.93	2.60	2.59	3.09	2.80
Dutch Reality TV	3.28	2.70	2.60	2.65	2.81
Italian Politics	3.15	4.01	4.01	3.30	3.62
Brazilian Politics	3.35	3.81	4.02	3.39	3.64
Dutch Politics	3.13	3.92	4.00	3.39	3.61
Nigerian Politics	3.33	3.98	4.14	3.33	3.69
Italian Football	3.12	3.91	4.00	3.37	3.60
Brazilian Football	3.10	3.90	3.87	3.25	3.53
Dutch Football	3.24	3.94	4.16	3.42	3.69
Nigerian Football	2.94	4.18	3.89	3.17	3.55
Italian Weather	3.28	3.88	3.97	3.42	3.64
Brazilian Weather	3.25	4.00	4.04	3.45	3.69
Dutch Weather	3.12	4.12	4.03	3.40	3.67
Nigerian Weather	3.24	4.02	4.33	3.37	3.74
Italian Generic	3.16	3.84	3.94	3.37	3.58
Brazilian Generic	3.27	3.85	3.92	3.38	3.60
Dutch Generic	3.19	3.84	3.93	3.48	3.61

3 SIGNING RELATIONSHIPS

The central theme around which this thesis is based is the Signed Ego Network Model (SENM). Of course, in order to obtain a signed version of the ENM, signed connections first need to be computed. Thus, this chapter aims to respond to the research question: how can a sign be computed for a relationship?

Precisely, the sections of this chapter propose a bottom-up signing methodology for signing relationships, then labels generated using this approach are compared using a variety of different models and, finally, the methodology is validated using known expectations of signed networks.

3.1 PROPOSED APPROACH

Given that relationships are made up of numerous interactions, it would be reasonable to approach the task of signing a relationship by first observing these fundamental building blocks. Indeed, this intuition is backed up by psychological research. It has been observed that the amount of negative interactions in a relationship can be a very reliable indicator of its health (Gottman, Markman, and Notarius, 1977). While a small amount of negative interactions is to be expected in any relationship, once the percentage of these negative interactions passes a certain tipping point, relationships will start to have numerous detrimental effects on those involved (Gottman, 1994). This threshold has been observed to be roughly 1 negative interaction in every 6, around 17%. Marriages eliciting more negative interactions than this are significantly more likely to end in divorce (Gottman et al., 1998) and parental relationships are more likely to see the child develop behavioural problems and perform poorly at school (Hart and Risley, 1995). Therefore, it seems appropriate to approach the signing of a relationship from the bottom up, first computing signs for each individual interaction and then using them to infer an overall sign for the relationship as a whole. See Figure 5 for an illustration of this threshold.

However, as the proposed signing method relies on a threshold, it may not be reliable for relationships with few interactions. Namely, given that the psychological observations found a 1:5 ratio, relationships with fewer than 6 interactions may not be reliable. For investigations using ENMs, this may be especially limiting in the outer circles, where there are fewer interactions. In response to this potential problem, the mean numbers of interactions are investigated. This is

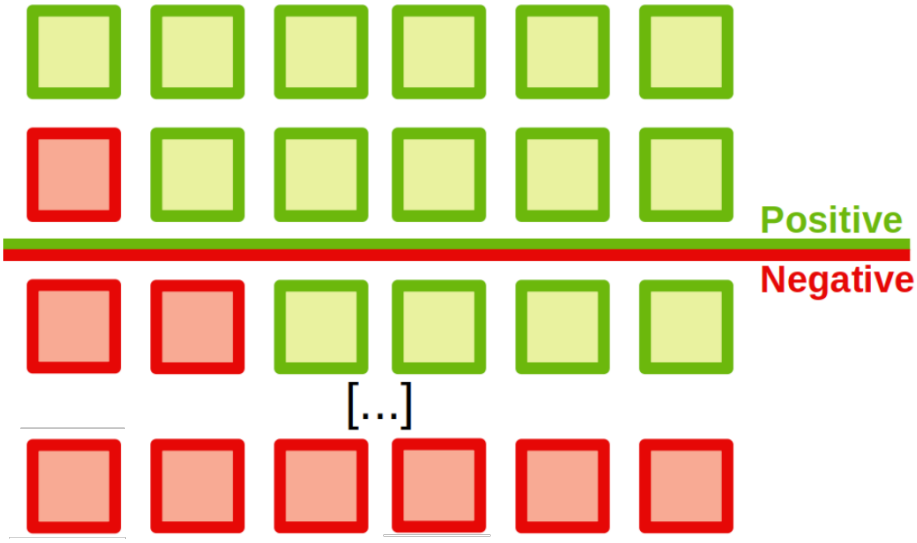

Figure 5: Illustration of the relationship signing threshold. Each row represents a relationship and each square represents an interaction. The relationships above the green/red line are labelled positive and those below negative

done for each Ego-Alter relationship and divided into each level of the ENM, in order to verify that we have enough data to properly apply the threshold. The results, summarised in Table 6, show that circles 1 to 4 have mean numbers of interactions that are equal to or greater than the required 6 for all the datasets utilised by this thesis. Only the outermost circle sometimes has too few interactions. This means that the signs of the first 4 circles can be considered to be reliable, while those of the final circle may sometimes need to be taken with a grain of salt.

Approaching the relationship signs by first looking at the individual interactions also allows some extremely powerful models to be leveraged. Sentiment analysis is the task of providing a sign (often "positive", "negative" or "neutral") based on the sentiment of a given text. As such, sentiment analysis can be applied to the individual interactions of relationships to obtain an estimate of whether it is a negative one. Sentiment analysis models have received huge amounts of attention over the past decade and can sometimes outperform individual human annotators (Hutto and Gilbert, 2014). What's more, there are many models (see Section 3.2) that have been specifically designed for use on short texts, making them perfect for use with X data and its 280-character limit.

Thus, the proposed relationship signing method consists of 2 main steps:

1. Compute a sign for each interaction in a relationship using sentiment analysis

Table 6: Mean number of interactions per Alter for each circle of the ENM

Dataset	Circle 1	Circle 2	Circle 3	Circle 4	Circle 5
American Journalists	59.09	23.66	12.69	6.69	3.18
Australian Journalists	83.90	21.80	12.56	6.33	3.02
British Journalists	49.14	20.10	12.26	6.60	3.13
Italian Journalists	202.24	80.74	37.20	16.45	6.28
Brazilian Journalists	118.79	52.97	24.42	12.78	5.01
Dutch Journalists	148.11	63.14	32.72	15.75	6.14
NYT Journalists	50.30	22.48	11.88	6.00	2.56
Science Writers	59.91	25.59	14.91	7.11	3.03
Monday Motivation	105.80	46.44	26.06	12.25	3.77
UK Users	86.53	44.34	22.53	10.81	3.72
British MPs	106.59	50.95	27.07	13.14	5.11
Baseline	174.19	68.69	34.95	16.71	6.48
Mediterranean	201.52	81.08	43.41	18.38	6.59
South America	193.80	80.25	36.37	16.26	5.58
Northern Europe	185.59	68.12	33.00	16.15	6.14
West Africa	140.65	67.69	34.26	17.54	6.14
Italian Reality TV	169.35	59.93	28.71	13.24	4.66
Brazilian Reality TV	132.50	57.44	30.27	14.19	4.81
Dutch Reality TV	156.63	81.65	41.20	19.49	7.22
Italian Politics	151.93	66.06	35.01	16.15	5.34
Brazilian Politics	153.33	74.37	37.26	17.11	5.61
Dutch Politics	141.42	64.67	33.48	16.70	6.00
Nigerian Politics	138.38	65.25	33.97	16.61	5.57
Italian Football	139.27	65.64	34.03	16.10	5.61
Brazilian Football	146.20	59.93	30.59	13.96	4.57
Dutch Football	169.77	79.33	42.46	20.17	7.31
Nigerian Football	141.97	56.50	37.05	15.84	5.65
Italian Weather	144.23	53.23	32.80	14.92	5.61
Brazilian Weather	126.08	55.05	27.53	14.08	5.20
Dutch Weather	175.14	80.26	41.25	22.84	7.55
Nigerian Weather	103.29	50.47	29.20	15.91	5.93
Italian Generic	157.92	67.29	35.31	16.85	6.23
Brazilian Generic	143.33	63.75	32.54	15.55	5.63
Dutch Generic	150.38	66.89	34.87	17.14	6.61

2. Use the psychology-based threshold of 17% to obtain a sign for the relationship as a whole

3.2 CHOICE OF MODEL

As stated in the previous section, there are plenty of well-performing sentiment analysis models. In order to gauge the susceptibility of the proposed relationship signing method to the choice of model, 4 models were chosen to generate a set of initial signs using the Baseline dataset. These signs were then used to investigate how much the different models agreed, as well as when they disagreed. These results are detailed in Section 3.3.

Recently, there has been a strong shift towards the use of transformer-based methods for Natural Language Processing (NLP). This is largely due to transformers' robustness and improved ability to process the sequential aspects of language. Reflecting this shift in focus, the models chosen for this study consist of a more traditional, lexicon- and rule-based model and 3 transformer-based models.

VADER

The first model is Valence Aware Dictionary and sEntiment Reasoner (VADER), a well-established sentiment analysis tool developed specifically for use with social media data (Hutto and Gilbert, 2014). VADER provides a compound sentiment score between -1 and 1 for a given text. This score can be converted into a positive label if it is above 0.05, negative if it is below -0.05 or neutral if it is between these values (Hutto and Gilbert, 2014). VADER was compared to 7 state-of-practice alternatives, as well as individual human annotators, using a test set of 4,200 tweets. It obtained an F1 score of 0.99, outperforming all other models and humans (Hutto and Gilbert, 2014).

BERTweet

The first BERT-based model used in this chapter is BERTweet (Nguyen, Vu, and Nguyen, 2020), a version of BERT (Devlin et al., 2018) that has been purposefully optimised for X data. Specifically, it was fine-tuned for the task of sentiment classification using a corpus of 850 million English Tweets collected between January 2012 and March 2020. BERTweet was tested using the SemEval 2017 (Task 4) corpus (Rosenthal, Farra, and Nakov, 2019), a common benchmark dataset for sentiment classification, which contains around 50,000 English Tweets; BERTweet achieved an F1 score of 0.73 (Nguyen, Vu, and Nguyen, 2020).

XLM-T

The next model is XLM-T (Barbieri, Anke, and Camacho-Collados, 2021), a fine-tuned version of XLM-RoBERTa (Conneau et al., 2019). This latter model is a general NLP model that was trained on 2.5TB of CommonCrawl data, containing 100 languages, which had been filtered following pre-established guidelines based on perplexity (Wenzek et al., 2019). The former was then further trained specifically for sentiment classification using 198 million Tweets from over 60 languages. XLM-T's performance varies from language to language but attained a mean F1 score of 0.69 when tested across monolingual datasets for 8 languages (Arabic, English, French, German, Hindi, Italian, Portuguese and Spanish). The F1 scores for 7 of these languages were between 0.69 and 0.78, however, Hindi only reached 0.56, highlighting the model's difficulty when dealing with certain languages. The English F1 score, 0.71, was obtained using a subset of 3,033 tweets from the SemEval 2017 dataset, thus, this model's performance seems to be similar to that of BERTweet.

BERT-C

The final model is a downstream version of BERTweet, also fine-tuned for sentiment classification, this time on a classified dataset. This model was released by HuggingFace (HuggingFace, 2022) and it is referred to here as the BERT Classified (BERT-C) model. Although no prior metrics for estimating the performance of this model are released, it is assumed that it will have a performance comparable to that of the original BERTweet model.

3.3 COMPARISON OF MODELS

Negative Percentages

First, the amount of negative interactions labelled by each model were compared. These can be seen as percentages in Figure 6. The models show a fair degree of variability, with around 30 to 45% for positive, 35 to 50% for neutral and 20 to 30% for negative. However, when looking at the relationship labels (Figure 7), there is a very tight percentage range for 3 of the models (VADER, BERTweet and BERT-C): between 60.71 and 63.53% positive (39.29% and 36.47% negative). By contrast, XLM-T shows almost equal numbers of positive and negative relationships (52.48% positive to 47.52% negative). Overall, these observations suggest that even though the models may have significant variations in their predicted labels for interactions, these differences shrink when it comes to labelling relationships. Thus, the proposed method of signing relationships appears to achieve very similar results regardless of the sentiment analysis model used for

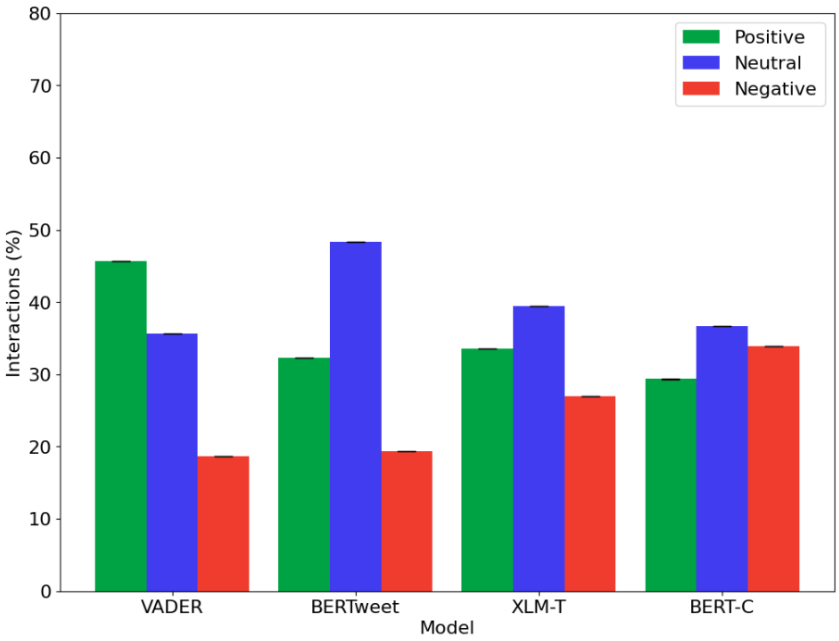

Figure 6: Percentages of positive, neutral and negative interaction labels estimated by each model (95% confidence intervals)

signing the individual interactions. Effectively, this robustness is due to the threshold-based nature of the relationship signing method, which can tolerate a certain degree of disagreement.

Note, as an additional remark, that the percentages in Figure 7 are more negative than the aforementioned observations of previous research (between 15.0% and 22.6% negative Leskovec, Huttenlocher, and Kleinberg, 2010b). However, as mentioned in Subsection 1, those results were observed in networks with publicly visible links, meaning that the number of negative links could have been suppressed due to the effects of Social Capital Coleman, 1988. On top of this, negative sentiment is known to spread far more easily on X than positive sentiment Schöne, Parkinson, and Goldenberg, 2021. So, it is expected that datasets without signs that are explicitly visible to users, especially those collected from X, will be more negative than these previous findings.

Agreement

Next, the level of agreement between each of the models was calculated using the proportion of predicted labels that matched exactly with the corresponding labels predicted by the other models. This was done to verify that the models are not just displaying similar amounts of negative relationships but that they are indeed agreeing on the signs of specific relationships. A matrix displaying these proportions for both the individual interactions labels and the relationships labels

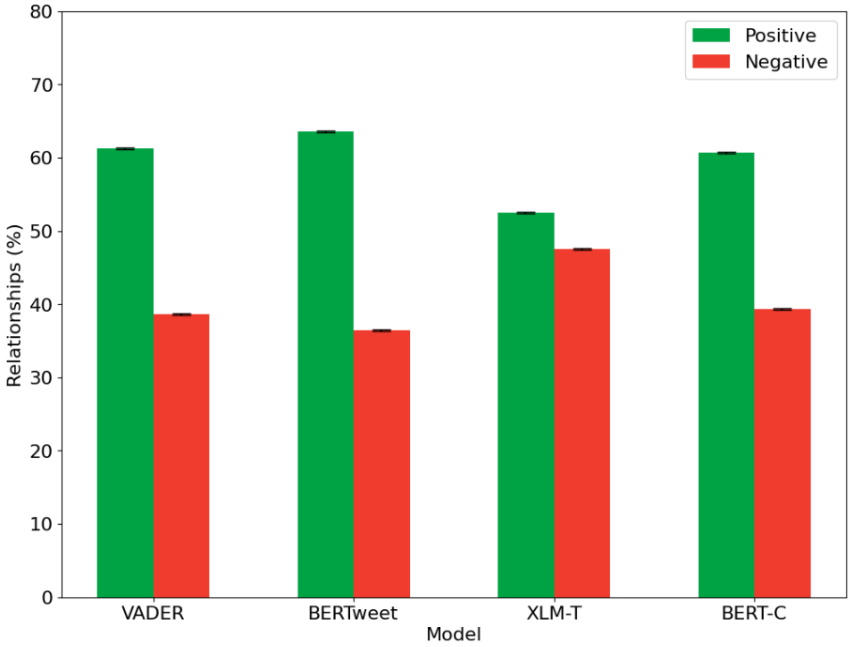

Figure 7: Percentages of positive and negative relationship labels estimated by each model (95% confidence intervals)

Table 7: The proportions of interactions that each pair of sentiment analysis models agree upon

	VADER	**BERTweet**	**XLM-T**	**BERT-C**	**Mean**
VADER	-	0.64	0.60	0.56	0.62
BERTweet	0.64	-	0.73	0.60	0.68
XLM-T	0.60	0.73	-	0.64	0.66
BERT-C	0.56	0.60	0.64	-	0.60

can be seen in Table 7 and Table 8, respectively. Due to the 1:5 (see Section 3.1) relationship signing ratio, only relationships with 6 or more interactions are included. Looking at the mean values of these results, the models agree for 60 to 68% of the interactions and 70 to 79% of the relationships. Not only do the models predict similar numbers of each label but they also agree on the vast majority of specific relationships. Interestingly, XLM-T, which predicted fewer positive relationships than the other models, has the highest mean agreement for relationships. This further illustrates that the relationship labels obtained are mostly independent of models and that, thus, the method of signing relationships proposed in this chapter can work irrespective of the model used to analyse the sentiments of individual interactions.

Table 8: The proportions of relationships that each pair of sentiment analysis models agree upon

	VADER	**BERTweet**	**XLM-T**	**BERT-C**	**Mean**
VADER	-	0.79	0.76	0.66	0.74
BERTweet	0.79	-	0.84	0.69	0.77
XLM-T	0.76	0.84	-	0.76	0.79
BERT-C	0.66	0.69	0.76	-	0.70

Disagreements

Finally, the relationships for which the models disagreed on a final label were investigated. This was done by comparing the percentage of negative interactions obtained from each model, thus, gaining a better understanding of the degree to which the models disagree. Again, only relationships with at least 6 interactions are included.

By plotting these negativity percentages for pairs of models, it is possible to visualise where the models are disagreeing, as in the example Figure 8[1].

However, given the fractional nature of these values, there are many points that overlap one another. To combat this, and to gain a more precise, numerical perspective, we then look at where the quantiles of these disagreements are. Specifically, for each relationship marked as negative when using model X (meaning that the fraction γ_X of negative interactions is above 0.17) and positive when using model Y (meaning that the fraction of negative interactions γ_Y is below 0.17), we compute the distribution of γ_Y. If our hypothesis is correct, we expect γ_Y to be concentrated in the area close to 0.17.

These quantiles can be seen in Figures[2] 9, 10, 11 and 12 for VADER, BERTweet, XLM-T and BERT-C, respectively. The exact values of the quantiles are also displayed in Table 9. These numbers show that the vast majority of disagreements are indeed happening in the area immediately above the 17% threshold. This suggests that, even when the models do disagree, they usually don't disagree by very much.

1 Observing the graphs, one may take note of the horizontal lines at the 0.0 mark on the y-axis. This corresponds to the case in which one model considers the relationship to be entirely positive but the other model still marks it as negative. While these strong disagreements are somewhat surprising, the majority of these occur before the 33% mark along the x-axes, i.e. close to the threshold, so most of them still correspond to relatively slight disagreements. What's more, the average number of interactions corresponding to these strong disagreements is 12.15, compared to 27.69 for all disagreements, meaning that strong disagreements are much more likely to happen for relationships with fewer interactions.

2 As some of the information in these plots is duplicated, for example, the comparison between model A and model B would be the mirror of the comparison between model B and model A, only the lower half of these plots have been included.

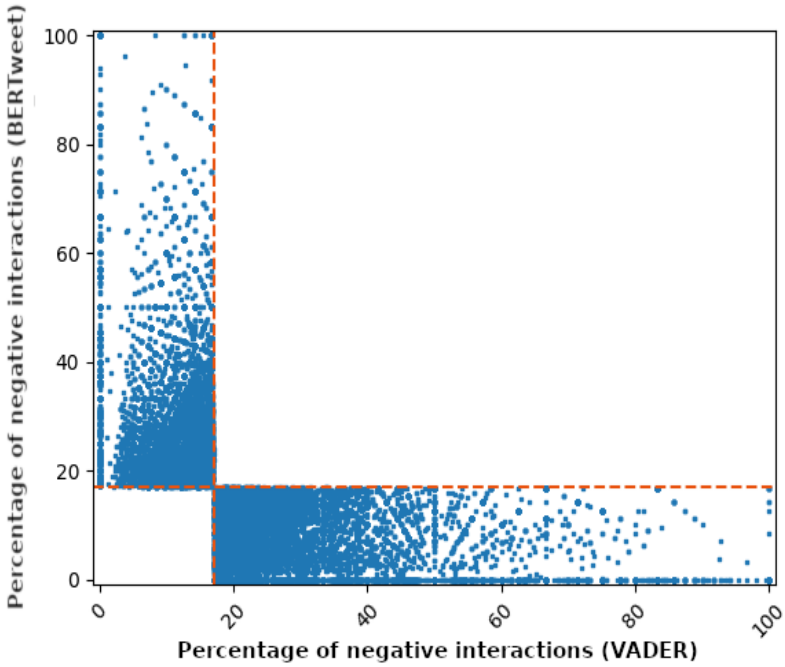

Figure 8: Example disagreement plot. Here, the relationships' proportions of negative interactions are denoted on the x-axis for VADER and on the y-axis for BERTweet.

Even the model that disagrees the most strongly with the others, BERT-C, has its third quantiles, i.e. 75% of its disagreements, under and around 40, which corresponds to approximately only 30% of the disagreement range $(17, 100)$.

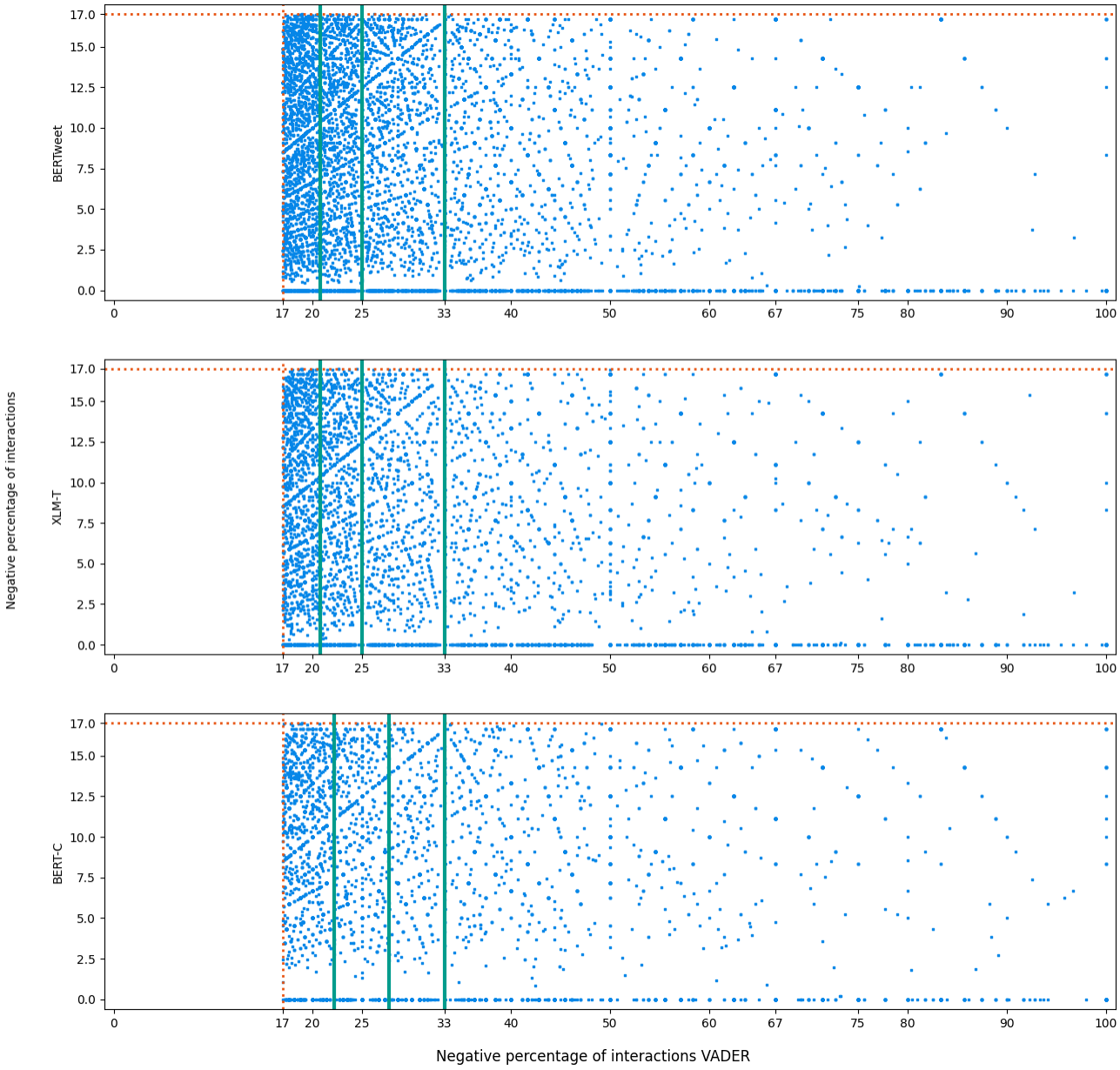

Figure 9: Disagreements between VADER and, from top to bottom, BERTweet, XLM-T and BERT-C. The vertical green lines denote quantiles of disagreements: 25%, 50% and 75%.

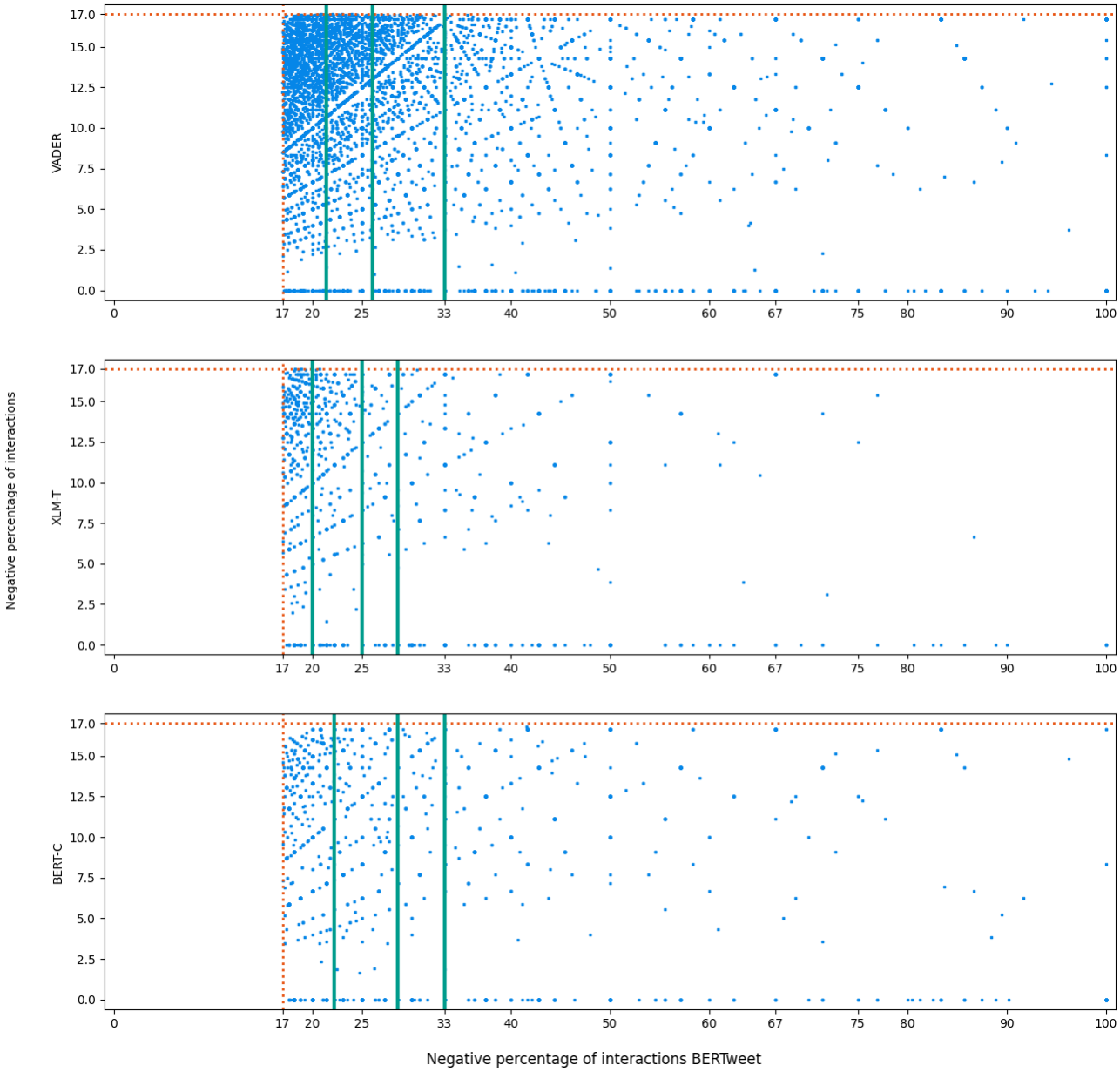

Figure 10: Disagreements between BERTweet and, from top to bottom, VADER, XLM-T and BERT-C. The vertical green lines denote quantiles of disagreements: 25%, 50% and 75%.

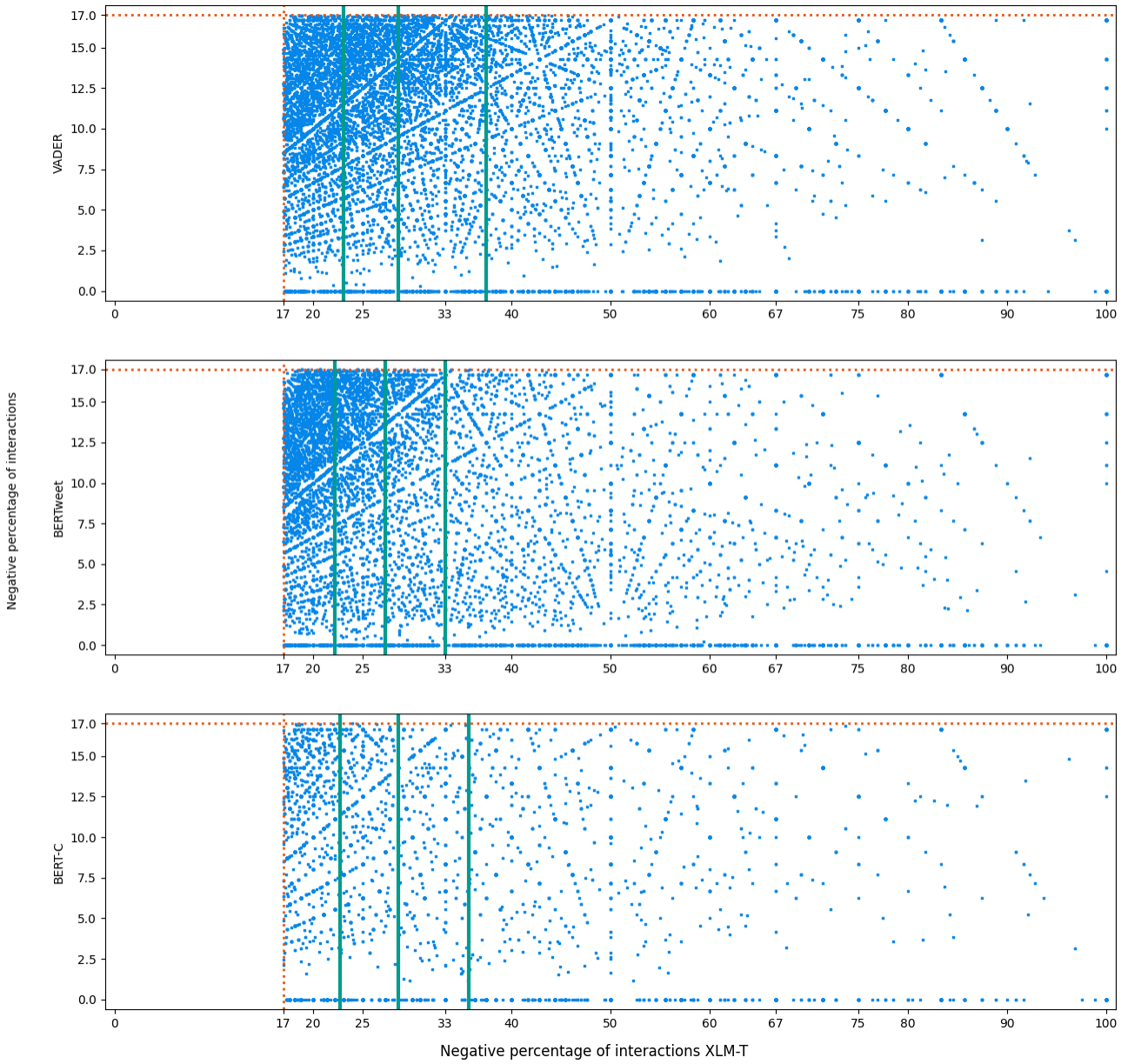

Figure 11: Disagreements between XLM-T and, from top to bottom, VADER, BERTweet and BERT-C. The vertical green lines denote quantiles of disagreements: 25%, 50% and 75%.

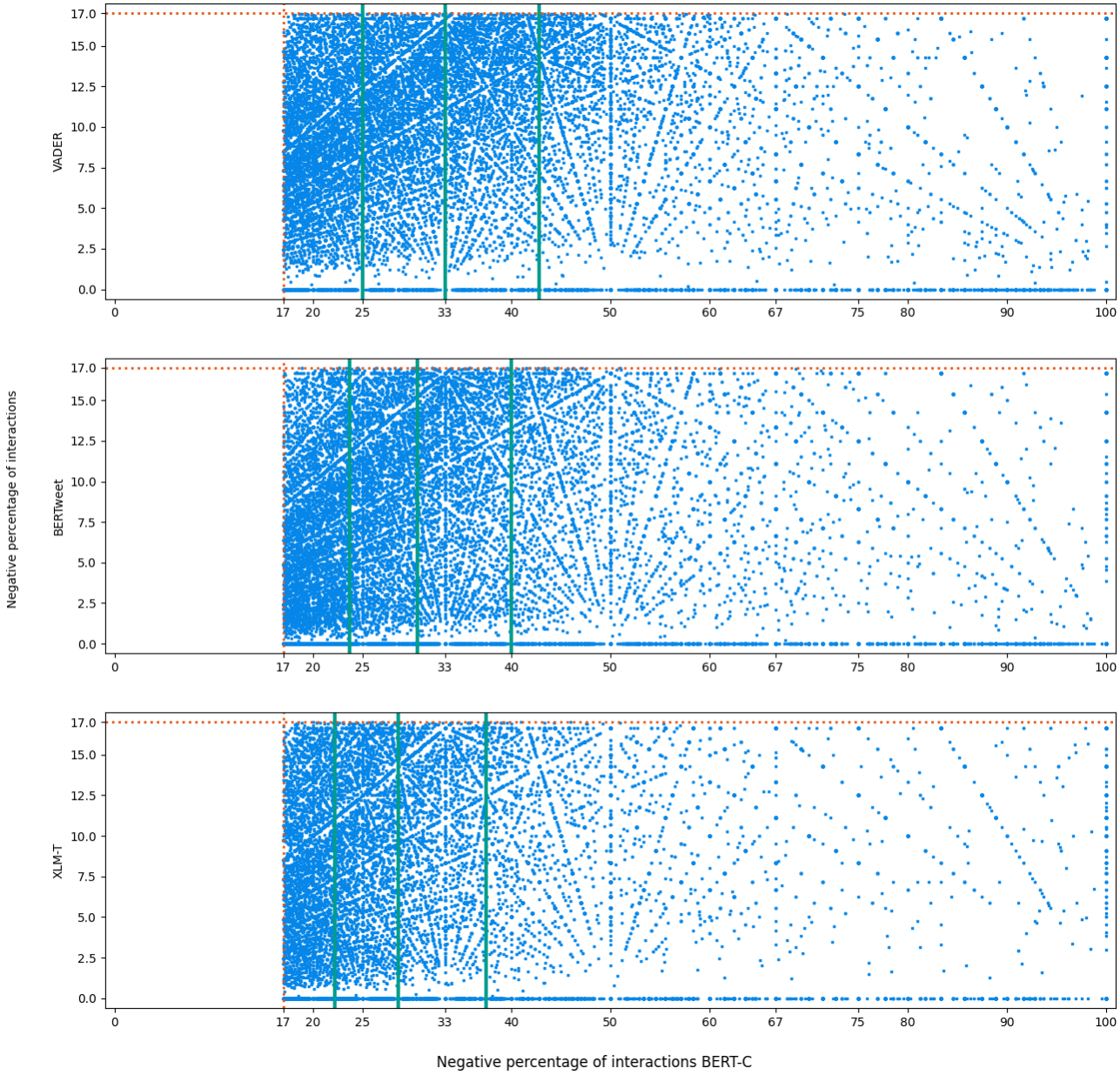

Figure 12: Disagreements between BERT-C and, from top to bottom, VADER, BERTweet and XLM-T. The vertical green lines denote quantiles of disagreements: 25%, 50% and 75%.

Table 9: Relationship label disagreement quantiles. The model giving a positive label is on the top and the model giving a negative label is on the left.

		VADER	**BERTweet**	**XLM-T**	**BERT-C**
Q1	VADER	-	20.83	20.83	22.22
	BERTweet	21.43	-	20.00	22.22
	XLM-T	23.08	22.22	-	22.73
	BERT-C	25.00	23.73	22.22	-
Q2	VADER	-	25.00	25.00	27.78
	BERTweet	26.09	-	25.00	28.57
	XLM-T	28.57	27.27	-	28.57
	BERT-C	33.33	30.52	28.57	-
Q3	VADER	-	33.33	33.33	33.33
	BERTweet	33.33	-	28.57	33.33
	XLM-T	37.50	33.33	-	35.71
	BERT-C	42.86	40.00	37.50	-

3.4 VALIDATION

As mentioned in Chapter 1, signed social networks are expected to follow certain patterns, such as those predicted by Balance Theory. Balance Theory defines expectations of the proportion of signed triads present within a network, based on observations of known signs (Heider, 1946). A triad is a group of 3 users who all have direct connections with one another. As a signed connection can be either positive or negative, the number of positive signs can range from 0 to 3: giving a total of 4 different signed triads. These are denoted as T_i, where i is the number of positive edges.

Here, these triad expectations are leveraged to validate the relationship signs obtained using the proposed method of signing relationships. In order to form the triads, an interconnected network of users is required. This is different from the standard data used for computing Ego Networks, where only the interactions between the Ego and the Alters are of interest. For triad analysis, Alter-to-Alter interactions are also required. Therefore, the Baseline dataset was used as it was the only dataset fitting this prerequisite at the time of investigation.

After computing all relationship signs, each triad was then tallied and the resulting numbers were compared to those of random chance in order to gauge how under- or overrepresented each triad is. Thus, observing whether or not the computed relationship signs match the expectations of weak Structural Balance Theory (i.e. an overrepresentation of T_3 and an underrepresentation of T_2). The weaker version is used here because, recalling from Chapter 1, it has been found to better fit online data. Specifically, the signs of the connections in the Baseline network were randomly shuffled. In this way, the number of positive and negative signs, as well as the structure of the Baseline's social network, remained the same, only the locations of the signs were changed. To improve the reliability of this validation, the random shuffling was repeated 10 times and the final results were compared using the mean values. This method of testing the expectations of Social Balance Theory has been previously established in the scientific literature (Leskovec, Huttenlocher, and Kleinberg, 2010b).

The further away the quantities observed in the real signed graph are from the random ones, the more "surprise" there is and the lower the likelihood of the computed signs occurring due to random chance. Here, *surprise* is defined as the number of standard deviations by which the observed number of Triad i differs from that of the randomly shuffled network. The precise formula used for calculating the level of surprise (taken from (Leskovec, Huttenlocher, and Kleinberg, 2010b)) is given in Equation 1. Here, $s(T_i)$ is the surprise for the observed number of Triad i, Δ is the total number of triads in the dataset, $p_0(T_i)$ is the a priori probability of T_i without the effects of Social Balance Theory (i.e. how many T_i would be expected in the network given

Table 10: Results of the triad analysis, with the counts and proportions of the observed triads from each model, along with the expected proportions (for a random distribution of signs) and the level of surprise.

Model	Triad T_i	Counts	Proportions	Expectation	Surprise
VADER	T_3	16,734	0.267	0.212	33.4
	T_2	19,018	0.303	0.431	-64.1
	T_1	16,934	0.270	0.287	-12.0
	T_0	10,020	0.160	0.064	94.9
BERTweet	T_3	21,439	0.342	0.232	65.5
	T_2	15,771	0.252	0.437	-93.8
	T_1	15,057	0.240	0.274	-18.8
	T_0	10,439	0.166	0.057	117.7
XLM-T	T_3	15,873	0.253	0.122	100.1
	T_2	12,715	0.203	0.372	-87.7
	T_1	15,946	0.254	0.377	-63.6
	T_0	18,172	0.290	0.128	120.9
BERT-C	T_3	20,683	0.330	0.222	64.8
	T_2	15,623	0.249	0.435	-93.8
	T_1	15,366	0.245	0.281	-20.2
	T_0	11,034	0.176	0.062	119.2

a random distribution of signs) and $\mathbb{E}[T_i]$ is the expected number of triads T_i. The counts and proportions of each triad, as well as their mean expectations and surprise levels, can be seen in Table 10.

$$s(T_i) = \frac{T_i - \mathbb{E}[T_i]}{\sqrt{\Delta p_0(T_i)(1 - p_0(T_i))}} \tag{1}$$

The main focus for this analysis is the surprise measurement (rightmost column), which indicates the number of standard deviations by which the predicted number of each triad differs from that of the randomly shuffled version. According to the weaker version of Structural Balance Theory, triad 3 should be overrepresented and triad 2 should be underrepresented, and this is indeed the case for all 4 of the models. This qualitatively confirms that the patterns of the extracted signs are compatible with what is observed in explicitly signed human social networks. Additionally, the surprisingly abundant T_0 provides an initial glimpse at the higher prevalence of negative relationships on X, which we explore further in the subsequent sections.

Before moving on, it is important to note that, quantitatively, this triad analysis does not provide a means of comparison between the models. In other words, the magnitude of the surprise in the expected direction (e.g., T_3 being overrepresented) is not a measure of how good the model is (because there is no such numerical notion of "correct amount of surprise").

3.5 CHAPTER SUMMARY

In this chapter, a novel method of signing relationships was proposed based on sentiment analysis of text-based interactions between users. Results using this methodology were compared using 4 different cutting-edge models and, while there were some differences for individual interactions, the 4 models appeared to agree for the majority of relationships. Finally, the computed signs of each model were validated against the expectations of Social Balance Theory. They were found to follow these expectations and to be significantly divergent from random change. Based on these results, the proposed signing methodology appears appropriate and reliable for the purpose of signing relationships on X.

In the following chapters, these signs are used to generate SENMs, which are in turn investigated, applied to a networking task and studied across various cultural and online communities. In the interest of time, all subsequent analyses are conducted using only the signs of VADER. As the previous comparisons of sentiment analysis models and triad analysis only serve as validation layers, they do not provide a clear idea of which model is most suited to the task of signing relationships. Therefore, VADER was chosen for the English-language datasets as it is lightweight and known to annotate individual Tweets more accurately than individual humans Hutto and Gilbert, 2014. While, XLM-T was chosen for the non-English datasets (see Sections 2.2 and 2.3).

4

PROPERTIES OF THE SENM

Following on from the validation of the relationship signing methodology proposed by this thesis, signs can now be reliably applied to the ENM to obtain the SENM. This chapter conducts an in-depth study of the properties of this novel extension. First, differences between the levels of negativities in users' full and active networks are observed, followed by a circle-by-circle analysis of each layer of the SENM, as well as how these observations differ for different types of users. Finally, as the original ENM is based on the cornerstone of cognitive capacity, the impact of negative relationships on cognitive load is investigated.

4.1 FULL AND ACTIVE NETWORKS

Recalling from Section 1.2, an Ego's full network includes every interaction they have ever made, even if they only communicate with that person once in their lifetime. Unfortunately, given the number and size of many of the datasets, coupled with the computational complexity of sentiment analysis, it was not feasible to generate signs for the full networks of every dataset. Therefore, while signs were generated for the active networks of all the datasets, only datasets collected before February 2023 had signs computed for their full networks. Additionally, due to the differences in negative relationship rates between the monolingual and multilingual models (discussed in Section 3.2), only the English language datasets are included here to ensure a reasonable comparison. Thus, the percentages of negative relationships in the full and active Ego Networks of 9 of the datasets are displayed in Figure 13. The exact numbers of these percentages as well as those of the remaining active and, where possible, full networks are also available in Appendix B.1.

First, focusing on the full networks, the datasets display levels of negativity within and slightly above the previously observed range of 15.0% to 22.6% (Leskovec, Huttenlocher, and Kleinberg, 2010b) (mentioned in Subsection 1.3). Specifically, the full Ego Network negativities all fall between 16.45% (Monday Motivation) and 31.58% (NYT Journalists). Given that the signs of the links in the current datasets are not explicitly visible to the users, and that, therefore, social pressure towards having positive links will likely be reduced, these observations are very much in line with a priori expectations.

By contrast, the active networks show significantly higher, albeit more varied, levels of negativity, between 21.83% (Monday Motivation)

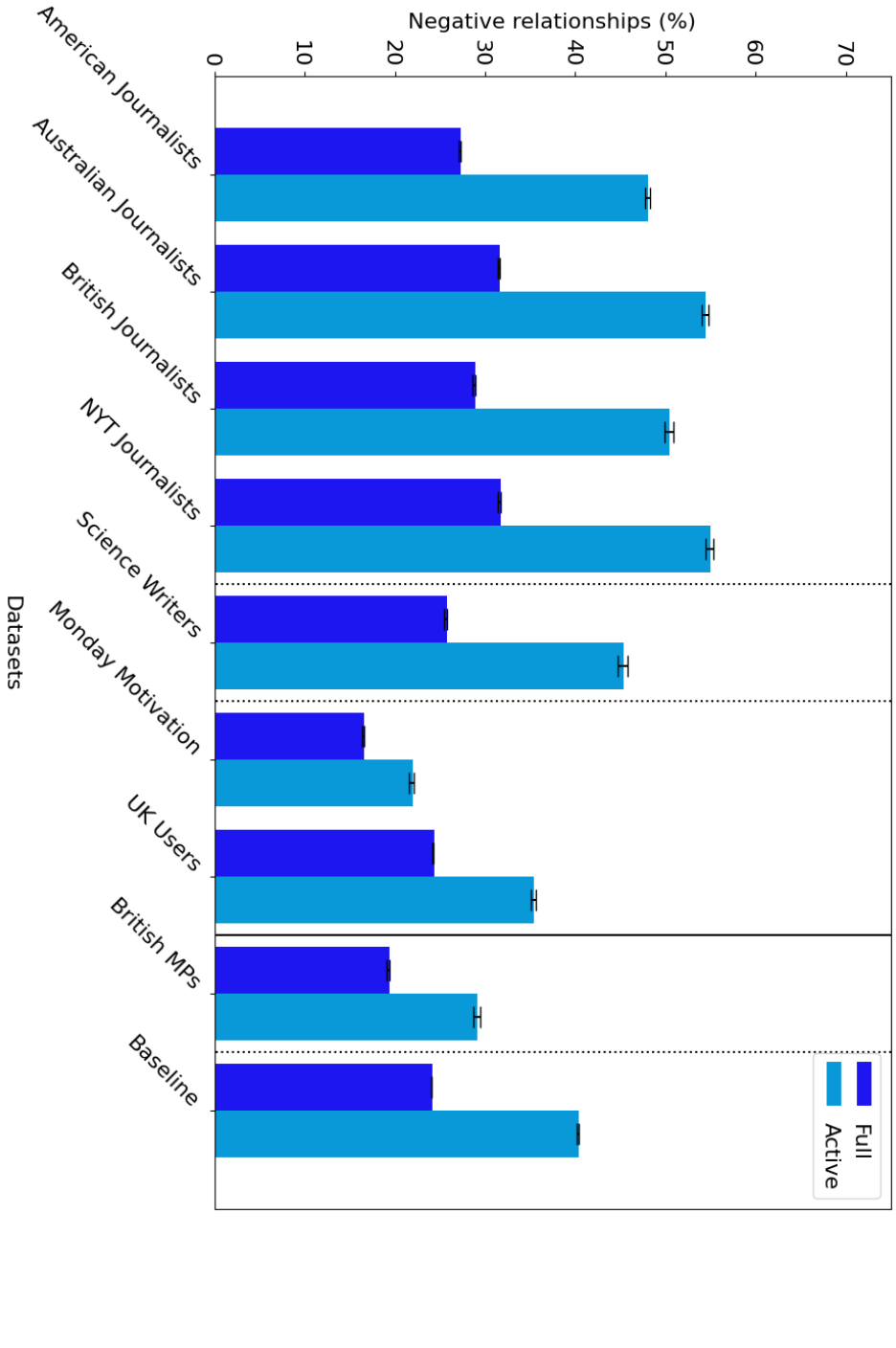

Figure 13: Percentages of relationships that are negative for the full and active networks of each dataset (95% confidence intervals)

and 54.89% (NYT Journalists). This increase in negativity from the full to active networks suggests that individuals have proportionally greater numbers of negative relationships amongst close contacts with whom they engage frequently than amongst acquaintances. Messages containing or eliciting negative emotions have previously been shown to elicit stronger responses (Baumeister et al., 2001) and to spread faster (Rozin and Royzman, 2001) than positive ones. Therefore, one explanation for the higher negativities of the active networks could be that, because the users of the active networks are communicating more frequently, any negative content that enters a user's Ego Network is more likely to be dispersed along the more active connections. Therefore, the connections of the active networks may display higher negatives because they have an elevated risk of being exposed to and spreading negativity. Thus, the more engaged an individual is, the seemingly greater the likelihood their relationships have of being negative.

The higher negative percentages of the active networks compared to the full networks suggest that Egos tend to have proportionally more negative relationships with Alters they engage with frequently than with acquaintances. It has previously been found that negative emotions and communications elicit a stronger response than positive ones (Baumeister et al., 2001) and that negative entities appear to be more contagious (Rozin and Royzman, 2001). Thus, one explanation for why the active networks are more negative is that, because the active users are spending more time engaged in online platforms, they have an elevated risk of being exposed to negativity, and therefore are more likely to be both subject to and to spread negativity. This is further supported by the observation that the journalist datasets are the most negative as it has previously been hypothesised that journalists are likely to be more cognitively involved with X than other types of users, such as politicians and generic users (Toprak et al., 2021). Therefore, the effects of exposure to negativity may affect them more strongly, which would explain the bigger relative increase in the negativity of their active networks. Conversely, this would also explain why the British MPs dataset had the lowest rate of negative relationships out of all the specialised networks, as they are thought to be generally less engaged with X than journalists (Toprak et al., 2022b).

In addition, although the increase in negativity from full to active network is most pronounced for the journalist datasets and science writers, this change is observable for all 9 of the included datasets. Therefore, rather than being a unique feature of any specific community, it appears that increased negativity is an inevitable byproduct of engaging with X. Investigating whether this phenomenon is observable for other social platforms, as well as how the effects differ, could be an interesting avenue for future research.

Next, as can be seen in Figure 13, most of the specialised users display higher percentages of negative relationships, compared to the generic users. However, this difference is fairly small for the full networks, with Baseline and the generic UK Users dataset actually containing more negative relationships (24.05% and 24.22% respectively) than the British MPs (19.24%) and nearly as much as the Science Writers (25.62%). By comparison, the difference between the active networks is much starker. With the only exception of the British MPs (whose change in negativity better matches those of the generic datasets), the least negative specialised dataset, Science Writers (45.23%), was nearly 5 percentage points more negative than the most negative generic dataset, Baseline (40.31%).

The greater negativities of specialised users also support the hypothesis that more engaged users are more likely to have a greater number of negative relationships, as mentioned in the previous subsection. On top of this, it has previously been hypothesised that journalists are likely to be more cognitively involved with X than other types of users, such as politicians and generic users (Toprak et al., 2021), which would explain both the higher negativities of the journalist datasets (between 47.97% (American Journalists) and 54.89% (NYT Journalists)) and the surprisingly low negativity of the British MPs dataset.

The active networks of the different cultural datasets are not explicitly discussed here. However, there is an in-depth comparison of the differences between these groups in Section 6.1, which is specifically dedicated to analysing cultural differences.

4.2 CIRCLE-BY-CIRCLE ANALYSIS

Next, the negativities across each circle of the SENMs are analysed. As mentioned in Section 2.5, in order to standardise the results, circle analyses focus on Egos with exactly 5 circles. The mean negativities of each circle can be seen in Table 11, where the circle with the highest percentage of negative relationships for each dataset is displayed in bold. The proportions of negative relationships are found to be disproportionately higher at the innermost circles of the ENM, decreasing steadily towards the outer layers. Specifically, the most negative circle is circle 1 for 11 of the datasets, circle 2 for 15 of the datasets and circle 3 for 7. This is very surprising as the inner sections of the ENM are usually associated with an individual's most trusted and similar connections. Indeed, as mentioned in Section 1.2, one of the four components from Granovetter's definition of tie strength is reciprocal services Granovetter, 1973, and reciprocity is thought to be very closely related to trust Ostrom, 2003. What makes these findings even more surprising is that the aforementioned effect of social capital (see Section 1.3), which creates a bias towards maintaining positive connec-

tions, would be strongest in the innermost circles, where individuals are expected to be the most tightly knit.

Table 11: Mean number of negative relationships for Egos with 5 circles for each circle, with percentages in parentheses and the most negative circle of each dataset in bold. The final column contains the difference in percentage points between the least and most negative circles of each dataset.

Dataset	C_1	C_2	C_3	C_4	C_5	$\text{Range}_{C_1C_5}$
American Journalists	**0.99 (61.37%)**	3.15 (59.13%)	8.53 (56.85%)	22.28 (53.33%)	60.23 (47.32%)	14.05
Australian Journalists	**1.09 (77.30%)**	3.34 (70.14%)	9.08 (66.74%)	25.27 (62.82%)	73.03 (54.21%)	23.08
British Journalists	**1.16 (63.33%)**	3.63 (57.98%)	9.76 (57.85%)	27.02 (56.22%)	70.94 (49.77%)	13.56
Italian Journalists	**1.12 (89.47%)**	3.57 (87.36%)	10.59 (81.85%)	33.14 (77.87%)	120.10 (70.61%)	18.86
Brazilian Journalists	1.78 (64.04%)	5.90 (74.92%)	**15.66 (76.25%)**	41.62 (72.56%)	116.48(66.12%)	12.20
Dutch Journalists	**1.66 (71.60%)**	5.51 (70.74%)	15.54 (69.34%)	43.08 (65.30%)	122.69 (58.27%)	13.33
NYT Journalists	1.11 (67.21%)	**3.73 (68.66%)**	9.90 (67.05%)	24.73 (61.59%)	60.43 (52.70%)	15.96
Science Writers	0.82 (48.39%)	**2.90 (49.91%)**	7.97 (48.59%)	21.31 (48.11%)	55.87 (44.75%)	5.16
Monday Motivation	0.30 (17.72%)	0.97 (18.37%)	**2.46 (18.59%)**	5.79 (17.25%)	14.09 (13.58%)	5.01
UK Users	**0.64 (34.75%)**	2.00 (33.46%)	5.27 (33.54%)	13.14 (33.41%)	37.63 (32.81%)	1.93
British MPs	**0.58 (29.41%)**	1.88 (28.24%)	5.08 (28.07%)	13.09 (26.71%)	31.31 (21.33%)	8.08
Baseline	1.00 (56.25%)	**3.63 (58.84%)**	9.72 (57.64%)	24.28 (54.95%)	63.71 (50.60%)	8.24
Mediterranean	**1.25 (73.58%)**	4.06 (72.54%)	10.38 (70.77%)	27.07 (69.70%)	76.85 (63.82%)	9.75
South America	1.37 (76.42%)	**4.42 (76.76%)**	12.00 (76.38%)	28.71 (71.93%)	75.03 (63.43%)	13.33
Northern Europe	**1.26 (69.86%)**	3.94 (67.05%)	11.04 (64.48%)	27.45 (60.54%)	70.67 (53.89%)	15.97
West Africa	0.92 (55.40%)	**3.18 (56.81%)**	8.75 (55.94%)	21.25 (53.51%)	60.80 (51.17%)	5.64
Italian Reality TV	1.08 (66.15%)	3.85 (74.76%)	**10.58 (76.35%)**	26.38 (74.09%)	71.38 (68.86%)	10.20
Brazilian Reality TV	**1.31 (82.89%)**	3.83 (82.51%)	9.92 (82.07%)	24.31 (77.70%)	67.73 (70.09%)	12.80

Table 11: (continued)

Dataset	C_1	C_2	C_3	C_4	C_5	$Range_{C_1C_5}$
Dutch Reality TV	1.15 (71.00%)	3.97 (75.00%)	**10.97 (76.75%)**	27.73 (74.61%)	67.42 (68.36%)	8.39
Italian Generic	1.05 (67.78%)	**3.45 (70.40%)**	8.87 (69.50%)	21.59 (66.58%)	59.87 (62.18%)	8.22
Brazilian Generic	1.19 (72.65%)	**3.68 (73.72%)**	9.48 (73.04%)	23.43 (70.72%)	63.15 (64.33%)	9.39
Dutch Generic	0.93 (56.97%)	3.01 (59.12%)	**7.87 (59.39%)**	19.44 (57.59%)	51.82 (53.36%)	6.03
Italian Weather	1.25 (76.88%)	3.81 (76.66%)	**10.21 (79.68%)**	25.46 (78.87%)	70.66 (74.87%)	4.81
Brazilian Weather	1.18 (81.16%)	**3.65 (81.32%)**	8.88 (79.26%)	21.42 (77.27%)	56.26 (70.09%)	11.23
Dutch Weather	0.84 (62.67%)	**2.79 (65.00%)**	6.70 (64.32%)	16.20 (62.77%)	45.14 (59.43%)	5.57
Nigerian Weather	**1.01 (75.00%)**	3.10 (74.30%)	7.54 (72.66%)	16.74 (69.73%)	45.57 (63.96%)	11.04
Italian Football	1.46 (84.80%)	**4.85 (87.84%)**	12.39 (87.75%)	30.51 (86.40%)	83.41 (79.50%)	8.33
Brazilian Football	1.34 (82.28%)	**4.34 (82.81%)**	10.92 (81.31%)	27.13 (78.25%)	74.55 (69.89%)	12.92
Dutch Football	1.05 (67.13%)	**3.43 (70.90%)**	8.69 (70.84%)	20.35 (68.97%)	54.65 (63.44%)	5.57
Nigerian Football	1.20 (77.78%)	**4.12 (78.97%)**	8.95 (71.68%)	21.02 (65.55%)	59.51 (58.78%)	11.04
Italian Politics	1.56 (91.82%)	**4.95 (92.01%)**	12.25 (91.43%)	29.94 (89.70%)	84.98 (83.95%)	8.06
Brazilian Politics	1.54 (90.79%)	4.71 (92.79%)	**12.36 (92.88%)**	29.99 (90.56%)	80.57 (82.60%)	10.28
Dutch Politics	1.54 (85.14%)	**5.03 (87.13%)**	12.78 (86.78%)	31.09 (84.38%)	84.89 (78.17%)	8.95
Nigerian Politics	**1.16 (76.69%)**	3.47 (76.01%)	8.70 (75.84%)	20.57 (74.26%)	57.11 (68.62%)	8.07

Furthermore, negative percentages tend to be higher across all circles for datasets which are centred around a more specific topic. For example, the circle negativities of the Italian Generic Users are 67.78, 70.40, 69.50, 66.58 and 62.18 and each of these negativities is less than that of the corresponding circle for the Italian Weather dataset, 76.99, 76.66, 79.68, 78.87 and 74.87, which in turn are lower than the corresponding Italian Football negativities, 84.80, 87.84, 87.75, 86.40, 79.50, and so on. Therefore, there is a clearly visible increase in negativity for more specific topics and this increase is observable for all cultures and at all levels of the SENM.

Comparing the proportions of negative relationships between the different types of users, there appears, once again, to be a divide between specialised and generic users. This difference becomes even more noticeable when the journalists are compared to the non-journalists. Indeed, the variations in negativity across the circles appear to be much greater for journalists than for any of the other datasets. The most stable journalist dataset, the British Journalists, drops by 13.56 percentage points from the most negative to the least negative circle. By contrast, the biggest variation for the non-journalists is the Generic Brazilian dataset with 9.39 percentage points. These observations lend further support to the notion that increased levels of engagement with X lead to increased levels of negativity. Egos engage the most with their innermost circles and this is where the strongest concentration of negative relationships is found. What's more, the difference between the negativity at this innermost level and that of the outer level is greatest for the users expected to be the most engaged category of users (i.e. the journalists). Otherwise said, the most negativity is found at the highest levels of engagement and this is true at every level of the Ego Networks as well as between different types of users. This could also explain why the T_0 triads (triads with no positive edges) in Section 3.4 were so prevalent.

Despite these observed differences in the proportions of negative relations across the circles, an observable ratio similar to that of the circle sizes appears to be fairly consistent across all datasets. As can be seen in Table 12, the lowest mean scaling ratio is 2.72 (British MPs) and the highest is 3.89 (Nigerian Weather), making this scaling ratio very similar to the 3 of the base ENM.

Table 12: Scaling ratios of negative relationships counts between circle sizes

Dataset	Circle 1-2	Circle 2-3	Circle 3-4	Circle 4-5	Mean
American Journalists	3.18	2.71	2.61	2.70	2.80
Australian Journalists	3.06	2.72	2.78	2.89	2.86
British Journalists	3.14	2.68	2.77	2.63	2.80
Italian Journalists	3.12	2.78	2.98	3.29	3.04
Brazilian Journalists	3.88	2.70	2.53	2.55	2.91
Dutch Journalists	3.28	2.77	2.61	2.54	2.80
NYT Journalists	3.37	2.65	2.50	2.44	2.74
Science Writers	3.52	2.75	2.67	2.62	2.89
Monday Motivation	3.18	2.54	2.36	2.43	2.63
UK Users	3.12	2.64	2.49	2.86	2.78
British MPs	3.23	2.70	2.58	2.39	2.72
Baseline	3.62	2.68	2.50	2.62	2.86
Mediterranean	3.25	2.56	2.61	2.84	2.81
South America	3.22	2.71	2.39	2.61	2.74
Northern Europe	3.13	2.80	2.49	2.57	2.75
West Africa	3.47	2.75	2.43	2.86	2.88
Italian Reality TV	3.58	2.75	2.49	2.71	2.88
Brazilian Reality TV	2.92	2.59	2.45	2.79	2.69
Dutch Reality TV	3.46	2.76	2.53	2.43	2.80
Italian Politics	3.15	4.04	4.09	3.52	3.70
Brazilian Politics	3.28	3.81	4.12	3.72	3.73
Dutch Politics	3.06	3.94	4.11	3.66	3.69
Nigerian Politics	3.36	3.99	4.23	3.60	3.79
Italian Football	3.01	3.91	4.06	3.66	3.66
Brazilian Football	3.08	3.97	4.03	3.64	3.68
Dutch Football	3.06	3.95	4.27	3.72	3.75
Nigerian Football	2.90	4.60	4.26	3.53	3.82
Italian Weather	3.29	3.73	4.01	3.60	3.66
Brazilian Weather	3.24	4.11	4.15	3.81	3.83
Dutch Weather	3.01	4.16	4.13	3.59	3.72
Nigerian Weather	3.27	4.11	4.51	3.67	3.89
Italian Generic	3.04	3.89	4.11	3.61	3.66
Brazilian Generic	3.22	3.89	4.05	3.71	3.72
Dutch Generic	3.07	3.83	4.05	3.75	3.67

4.3 EFFECT OF NEGATIVITY ON COGNITIVE LOAD

Given the obvious differences in the effects that positive and negative interactions can have on a relationship, an additional investigation was conducted to examine whether interactions and relationships of differing sentiments exert different amounts of cognitive effort. As negative information is generally harder and more time-consuming for humans to process Baumeister et al., 2001, one would expect negative relationships to be more cognitively demanding than positive ones. Therefore, the hypothesis we tested is whether greater numbers of negative relationships are associated with smaller active Ego Networks. For this analysis, the mean active Ego Network sizes of users with an optimum number of circles equal to 5 were compared. The users' levels of negativity were measured using 3 different metrics and, due to the slight differences in negative percentages resulting from different models (discussed in Section 3.2), this analysis was only done for the English language datasets.

Before introducing the formal definitions of the 3 negativity metrics, let us denote $\mathcal{A}_i$ as the set of Alters in the active Ego network of Ego i. Considering the signs of the relationships with the Alters, $\mathcal{A}_i$ can also be split into $\mathcal{A}_i^+$ and $\mathcal{A}_i^-$, for Alters whose relationship with the Ego i is positive and negative, respectively. Further, let n_{ij}^+ and n_{ij}^- be the number of positive and negative interactions between Ego i and Alter j. Their sum is denoted as n_{ij}. Leveraging this notation, the first negativity metric l_1 corresponds to the proportion of negative relationships, i.e. the number of negative relationships that each Ego had, divided by their total number of relationships:

$$l_1(i) = \frac{|\mathcal{A}_i^-|}{|\mathcal{A}_i|}. \tag{2}$$

The second negativity metric measures the proportion of negative interactions, even if they belong to positive relationships, i.e. the number of negative interactions for each Ego, divided by their total number of interactions:

$$l_2(i) = \frac{\sum_{j \in \mathcal{A}_i} n_{i,j}^-}{\sum_{j \in \mathcal{A}_i} n_{i,j}}. \tag{3}$$

Finally, the third negativity metric follows the proportion of interactions that belong to negative relationships, even if the interaction itself is positive, i.e. the number of each Ego's interactions that correspond to a negative relationship, divided by their total number of interactions:

$$l_3(i) = \frac{\sum_{j \in \mathcal{A}_i^-} n_{i,j}}{\sum_{j \in \mathcal{A}_i} n_{i,j}}. \tag{4}$$

When compared against the Ego Network size, the first of these metrics directly investigates the cognitive effects of maintaining negative

relationships regardless of how often we interact with said negative contacts. The latter two metrics take a more fine-grained look at the role of interactions. Indeed, the second metric gauges whether negative interactions, rather than relationships, have a different impact on cognitive effort, even if the negative interaction is with someone we have a positive relationship with. The third metric checks whether interacting with negative relationships elicits a different level of cognitive effort, even if some of the interactions are positive.

The values of the metrics are defined between 0 and 1 (inclusive) and the Egos are grouped into bins based on their negativity values for each of the 3 negativity metrics. Given the inconsistent distributions of Egos between the datasets and negativity metrics, the Egos were binned into quantiles. This ensures that all the bins of a given dataset contain similar numbers of Egos, although it does mean that the bin boundaries change between dataset and metric. The Egos' negativities metrics are compared to the sizes of the users' active Ego Networks and the number of users' interactions: both statistically and graphically.

For the statistical comparisons, Pearson's R was used. Our hypothesis is that an increase in negativity may correspond to an increase in cognitive effort, hence to smaller active Ego Networks and fewer interactions. This is based on our observations in Subsection 3.1 (see Table 6), which showed that specialised users, who show higher negativity levels, tend to display roughly half the number of interactions as generic users. Thus, a 1-tailed analysis was employed. The results showed no significant correlations for any of the datasets for either the active Ego Network sizes (p>.523 for all cases) or the number of interactions (p>.531 for all cases). This suggests that negativity does not decrease the size of Ego Networks, on average.

Next, binned boxplots were made to visualise the interplay between negativity and cognitive effort, for different classes of Ego negativity. We binned the Egos into quantiles, with respect to the negativity metrics, and then analysed the distributions of the active Ego Network sizes and the number of interactions in each bin. The corresponding boxplots for the 2 largest datasets in terms of Egos, Baseline and Monday Motivation, can be seen in Figures 14 and 15. The complete set of boxplots is available in Appendix C.1. Overall, the means, medians, boxes and whiskers of the boxplots are fairly flat across the bins (as expected given the non-significant correlations), suggesting that there is little or no effect of negativity on cognitive effort and that the differences observed in Subsection 3.1 between the numbers of interactions of the generic and specialised users are not due to the differences in negativity between these two classes of users.

These observations were followed up with t-tests conducted between pairs of bins for each dataset, with the null hypothesis being that there should not be any differences between them. This was done for both

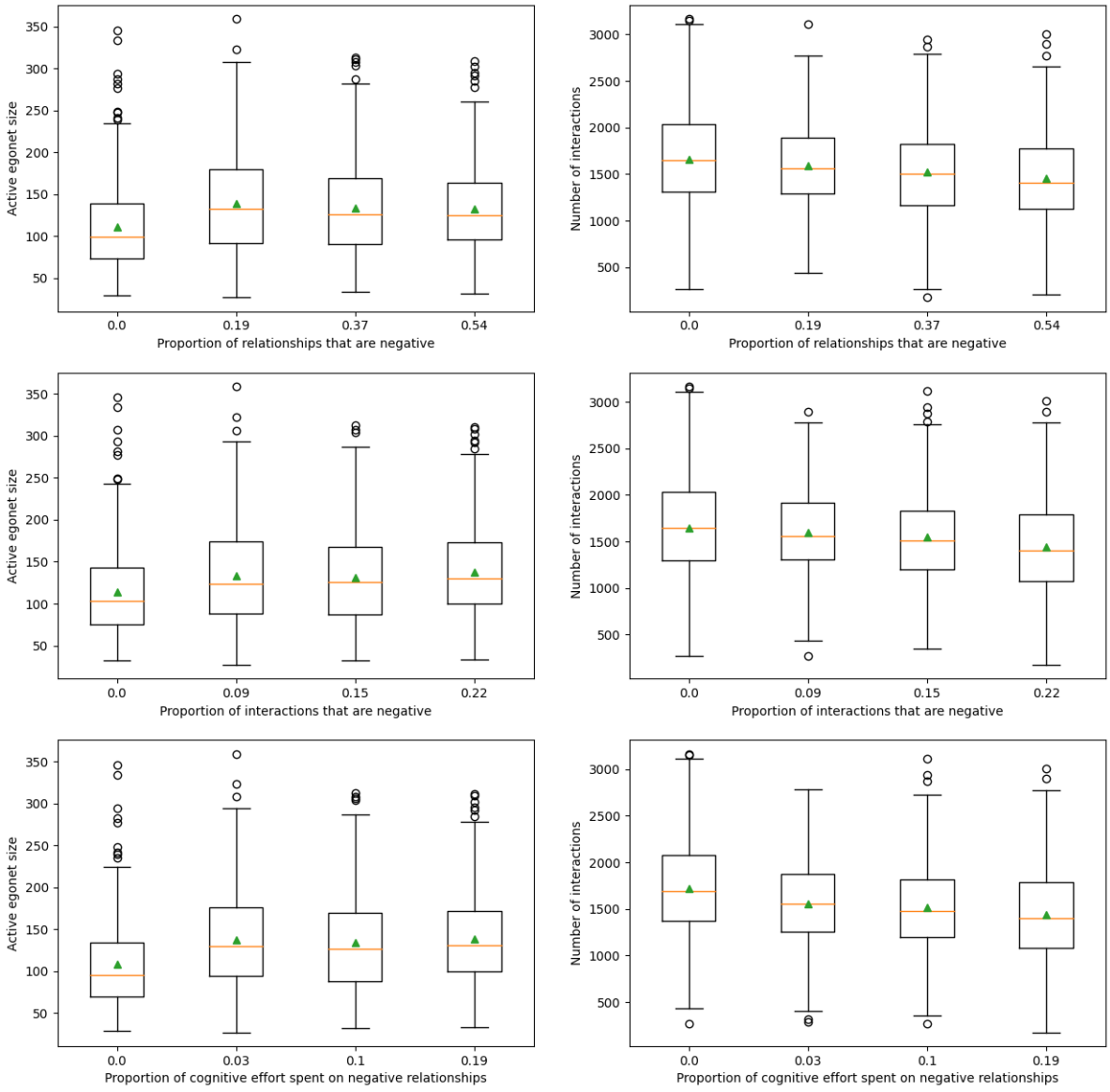

Figure 14: Boxplots for active Ego Network size (left column) and number of interactions (right column) against the 3 negativity metrics (top, middle and bottom) for the Baseline dataset. For each group of binned Egos, the boxplots display mean (orange line), median (green triangle), first to third quartile (box), 1.5 times the interquartile range beyond the box (whiskers) and outliers (black circles).

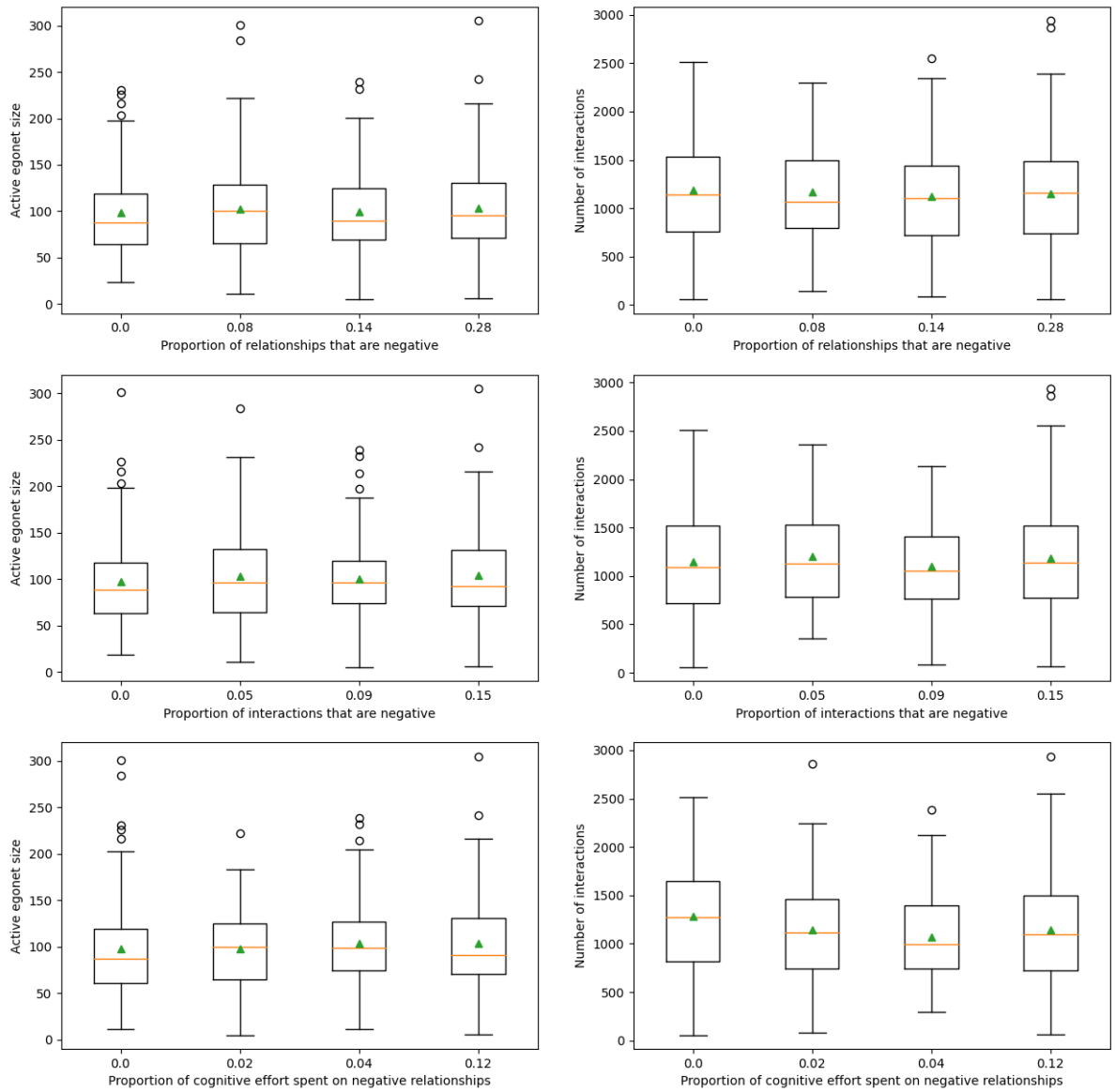

Figure 15: Boxplots for active Ego Network size (left column) and number of interactions (right column) against the 3 negativity metrics (top, middle and bottom) for the Monday Motivation dataset. For each group of binned Egos, the boxplots display mean (orange line), median (green triangle), first to third quartile (box), 1.5 times the interquartile range beyond the box (whiskers) and outliers (black circles).

the Ego network sizes and the number of interactions. The resulting p-values are displayed in Tables 13 and 14 respectively and the t-scores are available in Appendix D.1. The only dataset that displays consistently significant values is Baseline, which shows significant differences for all comparisons involving the first bin, for the Ego network sizes, and all comparisons involving the last bin for the number of interactions. The Baseline results would suggest that users with many positive relationships are likely to have slightly smaller Ego networks (i.e. fewer connections) and those with many negative relationships are likely to have more overall interactions. This could be a reflection of the hypothesis that the users who engage the most with the X platform are the most negative. Given that the Baseline dataset is significantly larger than the others this may be a weak effect that is only statistically significant when observing a very large sample size, however, as it is the sole dataset with these results, it is difficult to make any generalisable conclusions about the impact of negativity on cognitive effort.

These two results together would suggest that more positive users have fewer connections and interact less frequently overall but more intimately with the connections they do have (at least on the X platform). While more negative users have more, yet less intimate, connections with whom they interact less frequently compared to the positive users, they still end up interacting the most overall. However, while these results seem very promising, given some of the limitations of the negativity metrics analysis (i.e. the observations were only found to be significant for Baseline dataset), it would be pertinent to further investigate the interplay between these effects before making any further conclusions.

4.4 CHAPTER SUMMARY

The previous chapter proposed, performed and validated a novel method of signing relationships which could be applied to extend the ENM. This chapter has built upon this initial proposal by examining the properties of the resulting SENM. The proportions of negative relationships have been compared between full and active networks as well as across all levels of the SENM. The former investigation found that active relationships tend to be much more negative and the latter found that the inner circles of the SENM are also disproportionately negative. These two findings demonstrate that the relationships that X users engage with the most also tend to be the most negative. These findings were then followed up with a look at how the effects of negativity may affect cognitive load. Although there is some evidence of a weak influence of negativity on cognitive load, the results are not statistically significant enough to conclude anything definitively at

Table 13: The p-values from the pairwise comparisons between bins for Ego network sizes and negativity. Statistically significant values are displayed in bold.

	Dataset	Bins 1-2	Bins 1-3	Bins 1-4	Bins 2-3	Bins 2-4	Bins 3-4
Metric 1	American Journalists	0.743	0.152	**0.029**	0.294	0.076	0.431
	Australian Journalists	0.235	0.966	0.414	0.143	**0.024**	0.333
	British Journalists	0.162	0.354	0.381	**0.016**	0.643	0.072
	NYT Journalists	0.431	0.764	**0.002**	0.548	**0.001**	**0.000**
	Science Writers	0.850	0.384	0.773	0.447	0.913	0.514
	British MPs	0.127	0.189	0.904	0.610	0.151	0.230
	Monday Motivation	0.598	0.892	0.447	0.680	0.846	0.519
	UK Users	0.321	0.548	0.577	0.669	0.645	0.970
	Baseline	**0.000**	**0.000**	**0.000**	0.221	0.146	0.829
Metric 2	American Journalists	0.230	0.582	0.164	0.513	0.917	0.420
	Australian Journalists	0.944	0.825	0.384	0.712	0.315	0.155
	British Journalists	0.872	0.584	0.806	0.728	0.721	0.495
	NYT Journalists	0.289	0.375	0.236	0.787	0.967	0.770
	Science Writers	0.242	0.383	0.959	0.712	0.232	0.378
	British MPs	0.341	0.192	0.977	0.939	0.321	0.171
	Monday Motivation	0.388	0.633	0.314	0.658	0.881	0.550
	UK Users	0.326	0.089	0.387	0.558	0.875	0.426
	Baseline	**0.000**	**0.000**	**0.000**	0.667	0.439	0.212
Metric 3	American Journalists	0.323	0.791	0.274	0.511	0.995	0.473
	Australian Journalists	0.143	0.502	0.780	0.337	0.176	0.649
	British Journalists	0.180	0.295	0.793	0.747	0.199	0.298
	NYT Journalists	**0.045**	0.319	0.445	0.263	0.054	0.636
	Science Writers	0.142	**0.020**	0.768	0.483	0.207	**0.031**
	British MPs	0.169	0.101	0.540	0.895	0.377	0.321
	Monday Motivation	0.997	0.364	0.390	0.280	0.315	0.998
	UK Users	**0.036**	0.452	0.426	**0.002**	0.176	0.101
	Baseline	**0.000**	**0.000**	**0.000**	0.421	0.978	0.391

Table 14: The p-values from the pairwise comparisons between bins for number of interactions and negativity. Statistically significant values are displayed in bold.

	Dataset	Bins 1-2	Bins 1-3	Bins 1-4	Bins 2-3	Bins 2-4	Bins 3-4
Metric 1	American Journalists	0.175	0.910	0.414	0.148	0.560	0.359
	Australian Journalists	0.508	0.705	0.324	0.773	0.659	0.499
	British Journalists	0.257	0.902	0.692	0.200	0.450	0.595
	NYT Journalists	0.703	0.556	0.173	0.820	0.272	0.372
	Science Writers	0.659	0.719	0.318	0.930	0.617	0.545
	British MPs	0.726	0.239	0.385	0.373	0.579	0.718
	Monday Motivation	0.810	0.335	0.611	0.457	0.781	0.649
	UK Users	0.366	0.503	**0.012**	0.771	0.099	**0.042**
	Baseline	0.126	**0.003**	**0.000**	0.082	**0.001**	0.110
Metric 2	American Journalists	0.469	0.993	0.911	0.454	0.401	0.916
	Australian Journalists	0.567	0.606	0.100	0.929	0.223	0.184
	British Journalists	0.873	0.783	0.607	0.893	0.714	0.843
	NYT Journalists	0.271	0.087	0.597	0.528	0.491	0.163
	Science Writers	0.482	0.441	0.699	0.932	0.744	0.689
	British MPs	0.166	**0.030**	0.905	0.446	0.287	0.080
	Monday Motivation	0.504	0.475	0.668	0.127	0.821	0.226
	UK Users	0.462	0.993	0.466	0.448	0.129	0.440
	Baseline	0.270	**0.022**	**0.000**	0.166	**0.000**	**0.015**
Metric 3	American Journalists	0.114	0.089	0.661	0.950	0.269	0.228
	Australian Journalists	0.551	0.433	0.180	0.867	0.063	**0.041**
	British Journalists	0.422	0.643	0.350	0.780	0.097	0.194
	NYT Journalists	0.091	0.075	0.442	0.862	0.270	0.221
	Science Writers	0.760	0.380	0.855	0.442	0.569	0.240
	British MPs	0.368	0.099	0.806	0.437	0.565	0.196
	Monday Motivation	0.080	**0.003**	0.073	0.241	0.949	0.276
	UK Users	0.284	0.286	0.470	**0.021**	0.710	0.054
	Baseline	**0.000**	**0.000**	**0.000**	0.445	**0.010**	0.060

the current moment. However, this may be an interesting avenue for future research.

The remaining chapters in this thesis build upon the insights gained about the properties of the SENM by first testing an application and then by comparing differences in these, and similar, observations across different cultures and online communities.

5 APPLYING THE SENM TO STANCE DETECTION

Now that a reliable method for computing SENMs has been proposed and validated (Chapter 3) and its structure and properties have been investigated (Chapter 4), it would make sense to test how this information can be applied to existing research problems. Therefore, this chapter will introduce the task of Stance Detection (SD) and investigate how Ego Networks and Signed Ego Networks can be leveraged for it, comparing their performances against another cutting-edge model.

5.1 STANCE DETECTION

Essentially, SD aims to predict the stance of a given text towards a target entity (Biber and Finegan, 1988). Of course, being able to monitor the opinions of individuals or even overall trends in larger communities and populations can be extremely impactful. Especially, given its application to politics, where it can be used to quickly understand how people feel towards a given topic or to predict how they will vote. SD has often viewed interactions in isolation, predicting a user's opinion towards a given entity purely based on what they have written (e.g. Wei and Mao, 2019). However, recent research has shown that considering an individual's surrounding social network can greatly improve the accuracy of SD (Khiabani and Zubiaga, 2023), highlighting the influence of social connections on opinions.

Indeed, a model called CT-TN (which was introduced briefly in Chapter 1) combines predictions from a more traditional text-based model, RoBERTa (Liu et al., 2019), with multiple network features from the X social media platform: likes (a list of users whose posts have been liked by the target user), followers (the users who follow the target user) and friends (the users who are followed by the target user). This was done by creating an embedding for each feature and passing it through a classification model to obtain a prediction for each feature. The final prediction was then generated based on a majority vote of all the features (see Figure 16). The CT-TN model outperformed other competitive models, such as CrossNET (Xu et al., 2018) and TGA-Net (Allaway and McKeown, 2020) in six different experimental conditions.

However, the different network features that CT-TN requires are not always available. Indeed, the multiple different data sources required for CT-TN may be impossible or extremely costly to obtain in many situations. Therefore, it would be pertinent to investigate alternative approaches or features that are more parsimonious. Naturally, the

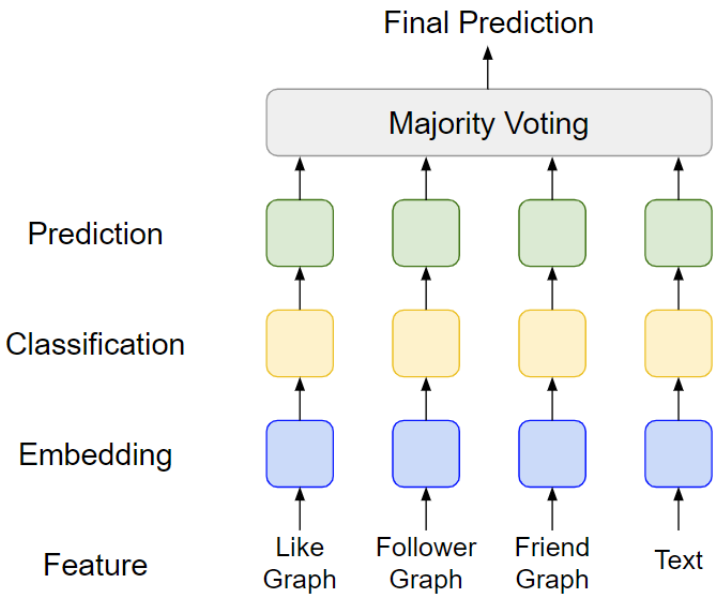

Figure 16: Architecture of the CT-TN model.

SENM and ENM seem like the perfect response to this limitation. The SENM only requires text-based interactions between users, which are often publicly available, and the unsigned version requires only the frequency of contact between users, allowing for implementation even when the contents of the interactions are not known or private.

Cross-Target Stance Detection

The objective of most SD studies is to classify target dependency in one of four main ways: Target-Specific, Multi-Related Targets, Target-Independent and Cross-Target (Alturayeif, Luqman, and Ahmed, 2023). Here, the primary focus is on Cross-Target Stance Detection (CTSD), which is when a model is trained using data for one target entity (source) and then tested on a different, although related, target entity (destination). For example, a model trained using texts containing opinions towards Joe Biden could be used to predict the stance of texts concerning another politician, such as Donald Trump or Bernie Sanders. This was the original task for which CT-TN was used.

Approaches to CTSD, as well as to SD more generally, differ based on the text's context and the particular relationship being discussed. On social media platforms such as X, there's often a focus on discerning the author's stance (supportive, opposing, or neutral) towards a specific proposition or target (Mohammad et al., 2016). Recent advances in SD encompass a range of linguistic features, such as word or character n-grams, dependency parse trees, and lexicons (Sun, Luo, and Chen, 2017). Moreover, there has been a shift towards end-to-end neural network methods that independently learn topics and opinions,

integrating them via mechanisms such as memory networks or bidirectional conditional Long Short-Term Memory models (LSTMs) (Augenstein et al., 2016).

Past studies have primarily relied on the text of a post to gauge its stance, neglecting the valuable insights that other features within social media platforms could offer. However, the performance of the aforementioned CT-TN model demonstrates the importance of considering structural features of the surrounding social network. Thus, knowledge about a user's connections can reveal important insights about the user themself and, therefore, about the texts they author.

5.2 FEATURES AND MODELS

In order to compute SD predictions, the data first need to be transformed into representations that are readable by a prediction model. For this, node2Vec (Grover and Leskovec, 2016) was applied to each of the previously established graph-based features (likes, followers, friends) as well as the novel Ego Networks and Signed Ego Networks, which, although they are converted into the same vector-space representation, can be better thought of as proxy measures of the way humans function socially. Node2vec is an unsupervised Deep Learning algorithm that uses a flexible, biased, random walk procedure to explore networks. The visited nodes can then be transformed into a vector space representation using a variety of methods, such as skip-grams or a continuous bag-of-words (Grover and Leskovec, 2016). This is similar to how the word2vec algorithm (Mikolov et al., 2013) treats words (nodes) and sentences (walks).

In addition to the graph-based features, text-based predictions were also used. These were generated using RoBERTa (Mikolov et al., 2013), an incredibly well-performing pretrained model that is used for many different natural language processing tasks. RoBERTa maps every token in a sentence to a vector representation in a continuous space.

The CT-TN model takes the predictions of each of these features, RoBERTa, likes, followers and friends, and obtains a final prediction using majority voting, where each feature's prediction acts as a vote for either "FAVOR" or "AGAINST". This allows for a thorough analysis of both textual and social network information, providing valuable insights for CTSD.

Aside from the aforementioned features, this paper also investigates two novel graph-based features: Ego Networks and Signed Ego Networks. These are also converted to a vector space representation using node2vec. Additionally, the unsigned Ego Network feature was also separated into inner (circles 1 and 2) and outer (circles 3 and beyond) circles, to better understand the importance of the different levels of the ENM for SD. Further details on how the Ego Networks are obtained are explained in Subsection 2.5.

CT-TN Model

The CT-TN architecture consists of three main parts. First, it uses text to understand and classify data. Second, it employs specialised components to handle followers, friends, and likes, using graphs to organise and analyse this social network data. These first two parts work together to make predictions. Finally, it combines all the predictions using majority voting method to give a complete result.

Model Hyperparameters

Alongside CT-TN, the base RoBERTa model was used for text embedding, trained with a batch size of 128, a dropout of 0.2, a learning rate of 3e-5 (AdamW), and 40 epochs. The graph embedding models were trained with the same dropout, a learning rate of 1e-2 (SGD), and 100 epochs. Each of the features was used to train a neural network model with two hidden layers for classification task. For the text-based embedding, this was done using RoBERTa, a batch size of 128, a dropout of 0.2, a learning rate of 3e-5 (AdamW), and 40 epochs. For the graph-based embeddings, this was done using node2vec, with the same batch size and dropout, a learning rate of 1e-2 (SGD), and 100 epochs.

Experimental Settings

The CT-TN model and the individual RoBERTa feature predictions were used as baselines against which to test the two Ego Network features. These were all prepared using few-shot cross-target training, whereby the training data consisted of roughly 1,000 source-target data points with 4 injections of destination target texts, increasing in size by increments of 100, from 100-shot to 400-shot (inclusive). For example, the Biden-Trump predictions were obtained by training on around 1,000 Biden texts with 100 Trump texts for the 100-shot condition, with 200 Trump tweets for the 200-shot condition, and so on. The stance predictions were then tested using between 500 and 800 data points (depending on the amount of remaining unseen data) that were solely related to the destination target (i.e. only using Trump-related texts for the aforementioned example). Each condition was computed 5 times with 5 different random seeds: 24, 524, 1024, 1524, and 2024. The final results were taken as the mean across the results of all seeds.

5.3 PERFORMANCE

Ego Networks

The performances of the CT-TN model, RoBERTa, and the two Ego Network features can be seen in Figure 17. Overall, they all perform

very well, with most reaching a macro F1 score of above 0.7 before 400-shot for all target pairs, with CT-TN sometimes even going above 0.8. However, RoBERTa does not perform quite as well as the others and only achieves macro F1 scores of around mid-0.6 for half of the target pairs.

Surprisingly, the signed and unsigned Ego Networks' F1 scores are very close, being within 0.01 of each other for 5 of the 6 target pairs (at 400-shot), and within 0.02 for the sixth pair (Sanders-Trump). This suggests that the additional information of signed connections does not provide a significant amount of information for the task of CTSD. Rather, it appears that the people with whom we interact regularly have an impact on our stances regardless of whether we have a negative or positive relationship with them.

The Ego Networks appear to perform slightly worse than the CT-TN model. However, as they only require interaction data, they could be used as a viable alternative whenever specific network features are not provided or obtainable for a given dataset. Moreover, as the signed and unsigned Ego Networks achieved similar performances, one could focus on employing the unsigned version, which would require even less data: only the frequencies of interactions, without the need for their texts.

Inner and Outer Circles

Next, observing the outer circles, one can see that they perform similarly to, and often even outperform, the full Ego Network. However, they are less consistent, as displayed by the Biden-Trump and Sanders-Trump target pairs. By comparison, the inner circles perform slightly worse overall, with performances closer to those of RoBERTa. Since the outer circles contain weaker social relationships, it seems that weaker, but more numerous, ties are more informative than stronger, but less numerous, ones when it comes to stance prediction. This is rather surprising given that previous research on a similar network-based task, link prediction, found that the more intimate inner circles are better predictors of where new relations will form (Toprak et al., 2023). Thus, there seems to be a disconnect between how people form new connections and how they are influenced by them. Indeed, paired with the fact that the signed and unsigned ENMs performed similarly, it appears that the existence of a social connection may influence an individual regardless of any qualitative aspects, such as closeness or polarity.

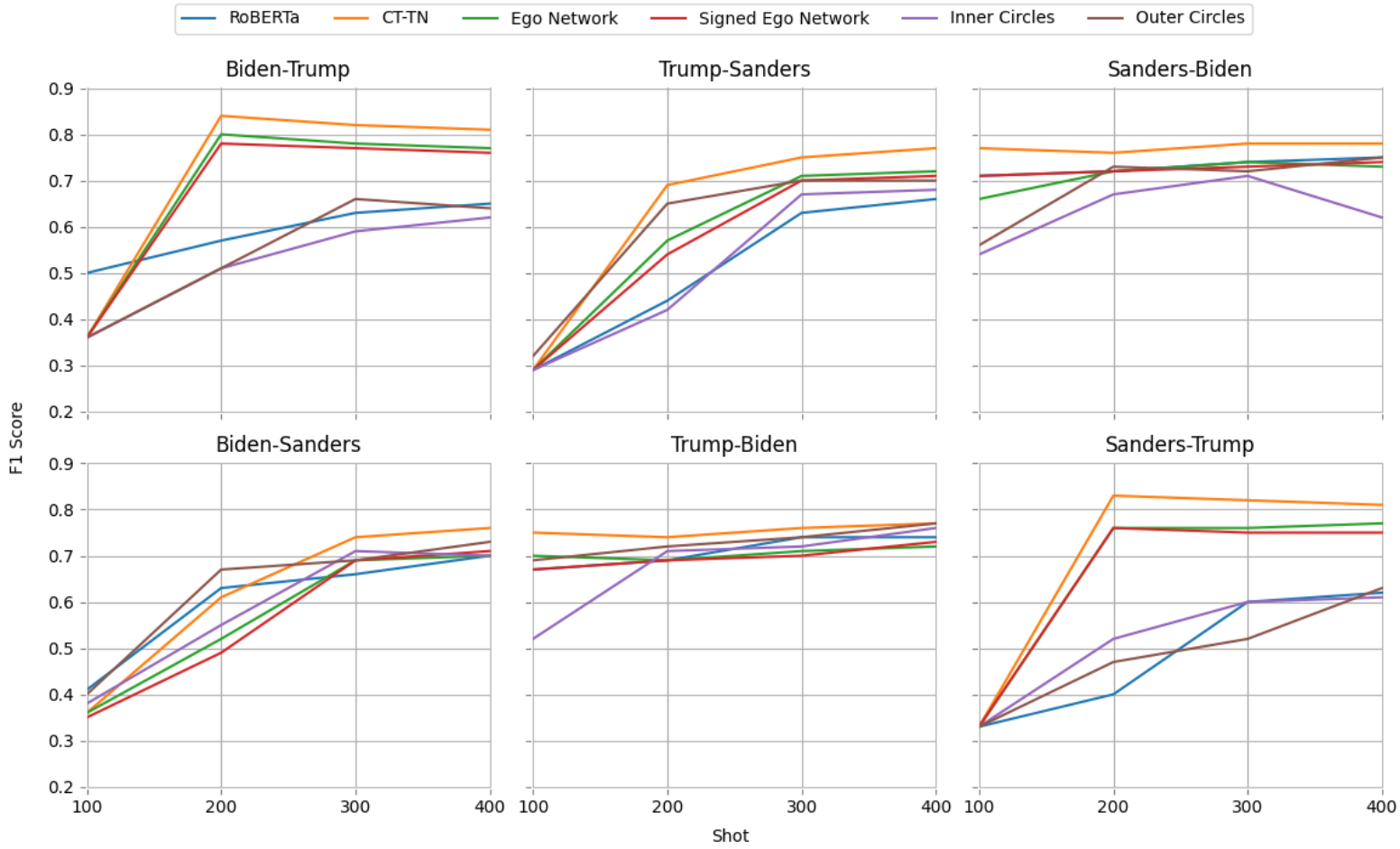

Figure 17: Graphs displaying the performances of the individual text-based RoBERTa, the CT-TN model and the signed and unsigned Ego Networks features, as well as the unsigned inner and outer circle features. The shot number is displayed along the X axis and the averaged macro F1 scores along the Y axis.

5.4 CHAPTER SUMMARY

As mentioned throughout this thesis, X has been one of the largest and most popular social network data sources for academic use for over a decade (Arnaboldi et al., 2013; Boldrini et al., 2018; Tacchi et al., 2022, 2024). Thus, the recent restrictions imposed on the accessibility of data from X, most notably the discontinuation of the Academic API, have presented a significant problem for many researchers who work with social networks and social media data. The results of this chapter help to partially bridge the gap caused by these restrictions. Concretely, while these features may perform slightly worse compared to more cutting-edge models, they are still competitive and present viable alternatives that could be used when not all the features required for CT-TN (or similar models) are obtainable.

Furthermore, by comparing the performances of the inner and outer circles, it appears that, while the inclusion of both leads to a more consistent performance across different target pairs, the outer circles on their own can often perform just as well and sometimes better. By contrast, the inner circles do not perform as well, suggesting that the greater number of less intimate connections in the outer circles are more important for predicting a user's stance.

6

DIFFERENCES ACROSS CULTURES AND COMMUNITIES

Now that the SENM and the methodology employed to compute it have been established, alongside an investigation of its practical uses, this chapter now explores potential differences that arise between different cultures and online communities. While the base ENM is known for being extraordinarily pervasive across many different societies, spanning many periods, there is no guarantee that the distribution of signs in the SENM is as consistent. Indeed, the omnipresence of the ENM structure is largely contributed to our innate cognitive constraints, however, as seen in Section 4.3, negativity does not seem to be strongly connected to cognitive load, hinting that the signs of the relationships we maintain may not be as inbuilt as their number.

Therefore, this chapter first observes differences in the percentage of negative relationships across communities and cultures. Then, in order to investigate whether these differences can be found for naturally forming social groups (i.e. in communities that were not specifically targeted for investigation), the most popular terms and topics negativities of subgroups within the 3 larger generic datasets, Italy, Brazil and the Netherlands, are investigated with regard to negativity.

6.1 NEGATIVITIES BETWEEN GROUPS

Given that the primary focus of this chapter is to investigate differences in the SENM between different groups, datasets are only included if they have one or more relevant counterparts from distinct groups. That is to say, the Journalist datasets (American, Australian, British, Italian, Brazilian and Dutch), larger geographical regions (Mediterranean, South America, Northern Europe and West Africa), Reality TV (Italian, Brazilian and Dutch), generic users (Italian, Brazilian and Dutch), weather (Italian, Brazilian, Dutch and Nigerian), football (Italian, Brazilian, Dutch and Nigerian) and politics (Italian, Brazilian, Dutch and Nigerian). The Baseline dataset is also included to provide a somewhat neutral point of comparison. The mean percentages of negative relationships in each dataset are organised into two tables: Table 15 for the pre-existing datasets and Table 16 for the novel ones. (Also, recall from Subsection 2.3, that the values of the datasets in the Geographical column of Table 15 are not just the countries listed in the Region column but also those of culturally-similar, neighbouring countries). The datasets in both tables are arranged into rows by region and into columns by type of user (Table 15) or topic (Table 16). Both rows

Table 15: Mean user negativities for datasets taken from previous papers, arranged by region and user type and ordered by negativity. Ranges between the Brazilian, Italian and Dutch datasets, ranges between all the datasets and ranges between all the topics are also displayed.

Region	**Geographical**	**Journalists**	**Reality TV**	**Range**
Brazil	65.67[1]	64.93	69.47	4.89
Italy	60.08[1]	63.87	64.97	4.54
Netherlands	54.66[1]	57.65	68.36	13.71
Nigeria	50.29[1]	–	–	–
Baseline	49.90	–	–	–
$\text{Range}_{\text{BR-IT-NL}}$	11.01	7.27	4.50	–
$\text{Range}_{\text{all}}$	15.78	7.27	4.50	–

[1]Note: The Geographical datasets are not exclusively the counties listed in the Region column, but also include a few neighbouring countries (see Subsection 2.3)

and columns are then sorted by negativity. When organised in this way, each negativity is larger than the one below it and to its left, with the only exceptions being Brazilian Journalists, Dutch Reality TV and Nigerian Weather. This strongly illustrates that both the geographical culture and the communities/topics that individuals are engaged with have a pronounced impact on the percentage of negative relationships that they maintain[1]. What's more, a clear pattern naturally emerges in the order of the topics in Table 16: from least controversial (Generic) to most controversial (politics). This strongly illustrates that the more a topic is controversial, the higher its negativity, and this holds across all cultures. Additionally, the negativities of the country-specific Generic Users are very close to those of the Geographical datasets, which also include users from neighbouring countries. This would suggest users from the countries that have been grouped together in the latter are fairly similar in terms of negative relationship percentage.

Next, the differences in the ranges of negativities are observed. The range columns in both tables show the range of negativity across different communities for a same country. Conversely, the range rows show the ranges across different countries for the same type of community. Note that, to better compare the values of the two tables, an additional range row was computed for only countries that have complete datasets in the new datasets, (i.e. all but Nigeria).

Looking first at the range rows of the pre-existing datasets, the cultural differences in negativity seem to be overpowered by the influence of the communities with which we are involved and that this effect is

1 It is worth noting that Italy and Brazil swap positions between the two tables. Suggesting that the order of geographical cultures based on negativity can vary slightly between user types and/or topics.

Table 16: Mean user negativities for datasets collected for this chapter, arranged by region and topic and ordered by negativity. Ranges between the Italian, Brazilian and Dutch datasets, ranges between all the datasets and ranges between all the topics are also displayed.

Region	**Generic**	**Weather**	**Football**	**Politics**	**Range**
Italy	66.75	74.24	77.96	84.34	10.10
Brazil	64.58	69.88	71.41	81.24	11.36
Netherlands	54.55	56.67	64.09	78.25	21.57
Nigeria	–	60.73	58.45	68.30	9.86
$\text{Range}_{\text{IT-BR-NL}}$	12.19	17.57	13.87	6.09	–
$\text{Range}_{\text{all}}$	12.19	17.57	19.51	16.03	–

stronger the more negative or controversial a social group or topic is. Indeed, this effect appears visible in the range rows of Table 15: the more negative types of users display lower ranges (i.e. less difference between the different cultures). However, this effect is not visible for the novel datasets, where the ranges seem to oscillate without any correlation with this expectation. Instead, the ranges appear to not be strictly determined by the topics. What's more, focusing on the ranges for the 3 countries that have available datasets in each novel category (i.e. excluding Nigeria), there is a topic with a lower range than the Generic dataset: Politics. This, of course, is the most controversial of all the datasets collected, which may indicate that cultural effects may only be suppressed by topics that are exceedingly polarising. In the datasets of Table 15, the same holds for Reality TV, which can also be considered as a topic eliciting quite controversial discussions. Thus, it appears that the impact of culture is not *always* being "overwhelmed" by that of topics or subcommunities with which an individual is engaged. Rather the strength of the influence a given topic or community has on an individual depends somewhat on their culture and that, although topics can overwhelm cultural differences in extreme cases (e.g. Politics). Intuitively, this conclusion seems logical as values and priorities can change dramatically between cultures, whereas politics can lead to strong and contradicting opinions in almost any culture. Indeed, the ranges could be taken as a measurement of how much the value of a given topic varies from culture to culture, rather than an inverse measure of how controversial it is. Therefore, future research may want to take steps to consider which topics could be considered controversial for each culture separately.

In contrast to those of the topics, the ranges of the countries (range columns) are more consistent: with all countries except the Netherlands displaying ranges that are relatively close to one another in both tables. This suggests that, while the impact a topic has on negativity

across a set of cultures can vary drastically, the cultural influence on negativity across a set of topics may be, at least somewhat, predictable.

6.2 MOST POPULAR TERMS

As previous works have found that communication behaviours can be starkly different between different types of users, such as between journalists and generic users (Toprak et al., 2021), as well as the differences in negativities already observed in this Chapter, an in-depth analysis of the most popular hashtags and words (this subsection), as well as topics (next subsection), within the datasets is conducted to see whether such effects can be observed. This is done for the generic cultural datasets as these generic datasets are less likely to be influenced by larger communities (as would be the case with any datasets already belonging to a specific community or group). Additionally, these datasets correspond to 3 of the 4 largest datasets included in this thesis and are therefore the most likely to have subcommunities within them.

First, the top 20 hashtags and the top 20 words were examined for both the full and active networks, for all Egos (not just those with an optimum circle number of 5). Both the hashtags and the words were standardised, removing diacritics, punctuation and capitalisation. Words were also removed if they were stopwords or if they were equal to or less than 4 letters in length, these latter 2 restrictions were not imposed on the hashtags. The top hashtags and words in the active networks were manually assigned 1 of 6 labels, corresponding to different topics. These labels were: "political" for politicians, governments, political topics or politic-only news channels, "COVID" for anything relating to the COVID-19 pandemic, "climate" for topics relating to green energy, renewable or the climate, "religious" for religious topics, "news" for general news services (i.e. not specifically political or religious) and "general" for everything else (these hashtags and their labels can be seen in Appendix E.1). Of course, natural language is often ambiguous and viewing individual terms in isolation can compound this problem. This means that a single word could be used to discuss different topics in different Tweets; when labelling for topics, the authors attempted to assign the label that would likely be the best match in the majority of uses.

Hashtags are generally used on X to indicate specific topics that are related to a Tweet, whereas the words are, by their nature, more general. Therefore, hashtags were chosen as the focus of this analysis. Using the percentage of negative relationships in the active networks, correlations were then calculated using Pearson's R for the number of times each topic appeared in the top 20, as well as the total and

proportional[2] number of times those hashtags were mentioned. For this test, there were 9 degrees of freedom. The only variable that had a significant correlation with negativity was the number of "general" topics in the top 20, $r(9) = -.64$, $p = .035$, meaning that there is a negative correlation between the number of "general" topics users frequently discuss and the number of negative relationships they have. One potential reason for this, given that individuals are more likely to engage in negative exchanges with someone whose beliefs differ from their own (Hutchens, Cicchirillo, and Hmielowski, 2015), is that higher levels of negativity are more likely to occur when the topic of conversation is specific, as strong opinions are more likely to have formed, as opposed to non-specific topics, where strong opinions are less likely to have formed. Indeed, this has previously been found for topics which have clear "sides", such as in politics (Coe, Kenski, and Rains, 2014). On top of this, being online means that you are more likely to be exposed to a greater range of opinions (Heatherly, Lu, and Lee, 2017) and those who spend more time on the platform (i.e. those who are the most engaged) are more likely to find someone they disagree with, thereby having an increased likelihood of negative exchanges. This could explain the higher observed negativities for journalists and TV watchers.

South America is surprisingly negative for a generic user dataset, even displaying slightly more negativity than the Brazilian Journalists. However, the top 3 hashtags in the active network of this dataset, as well as 8 out of the top 20, are news-related. All the other generic datasets have exactly 1 hashtag in the news category. This suggests that users in South America may be using X more as a news site than as a social media platform, which would explain why the generic users from this region are so similar to the journalists.

While the results of the analysis on the top 20 words revealed information that was very similar to that of the hashtags, one additional and rather interesting finding did arise: the usage of the word "Allah" in the West Africa dataset. There are many common phrases used in Muslim cultures that contain the word "Allah"; for example, "may Allah grant you health" or "may Allah strengthen the noble" (Rababah and Malkawi, 2012). Although there are differences in how these types of phrases are used between Muslin countries, they usually contain positive sentiment that is much stronger than the equivalent phrases used by other cultures; such as "hello" or "good morning". This could go some way towards explaining why West Africa is the most positive regional dataset and the second most positive overall. What's more, greetings in some West African cultures are known to be highly formalised and predominantly phatic (i.e. with the aim of establishing or

2 The proportional number being the how many times all the hashtags of each topic were mentioned in the dataset, divided by the total number of mentions of all hashtags in the top 20 for that dataset.

maintaining social relationships) (Goody, 1972). Thus, there could be a double effect towards positive communications. This specific use of certain phrases is a cultural difference that is relatively easy to detect when analysing social media data and could be directly affecting how individuals communicate, potentially enabling some very interesting insights about cultural differences to be revealed in future work.

6.3 TOPIC ANALYSIS

Following on from the investigation of the most popular hashtags and words, a further, Deep Learning-based approach was also conducted using a model called BERTopic model (Grootendorst, 2022). BERTopic is a topic modelling tool that uses a mixture of transformers and TF-IDF to identify and group important topics within a collection of natural-language documents. BERTopic was compared against 5 other state-of-the-art topic models in terms of two staple metrics: Topic Coherence (Lau, Newman, and Baldwin, 2014) and Topic Diversity (Dieng, Ruiz, and Blei, 2020). These measure how well a model's grouped terms fit with one another and how much variety there is among grouped words. BERTopic was found to consistently outperform the other models for Topic Coherence while also remaining very competitive for Topic Diversity (Grootendorst, 2022). To optimise BERTopic for multilingual data, the default transformer model it uses can be replaced with paraphrase-multilingual-MiniLM-L12-v2 (Reimers and Gurevych, 2019). This is a sentence-transformer model that is able to accurately process data in over 50 languages (Reimers and Gurevych, 2020).

Unfortunately, paraphrase-multilingual-MiniLM-L12-v2 can only take in the first 384 tokens from each document, making it impossible to pass in an entire User Tweet Timeline (which often contains a few thousand Tweets, each of up to 280 characters) as a single document. Tweets were therefore parsed individually, meaning that each Tweet was treated as being entirely distinct from all the others, even if they were created by the same user. An unfortunate consequence of this is that some of the topics may appear to be undeservedly important if, for example, they are being spam-tweeted by an individual user. In an effort to mitigate this, and in order to get a perspective of the topics being mentioned by all users in each dataset, the top 200 topics provided by BERTopic were collected and the number of distinct users involved with them was computed. The 20 topics with the largest numbers of unique users were then chosen as the focus of analysis. Effectively, this meant that each topic was counted a maximum of once per user, rather than once per Tweet related to the topic. These topics are listed and discussed in Subsection 6.3.

Using lists of key terms for each of the topics computed by BERTopic, it is possible to check each individual tweet to see if it is related to

any of the top topics. This led to a list of IDs of both the tweets and the users involved, which in turn allowed the topics to be matched to 2 negativity metrics. The first of these is the percentage of negative relationships within the Ego Network of users that tweeted in relation to each topic, taken as the mean of all the users. The resulting value provides a gauge of the negative impact a topic has on the relationships of those who are engaged with it. The second metric is simply the percentage of related tweets which are negative, which reveals how negative each topic is in isolation, i.e. irrespective of the surrounding network. These metrics are displayed in Table 17 (mean percentage of negative relationships) and Table 18 (percentage of negative tweets). In addition, some specific categories of topics have been chosen to be focused on in more detail. These categories represent a mix between the specific topics of the other datasets collected for this chapter (Politics, Football and Generic) as well as two additions: COVID, which provides a unique opportunity to analyse a single event which has had an impact on every single person across the globe, and Religion, which was noted as a specific topic of interest for future research in previous work (Tacchi et al., 2023). Both the aforementioned tables have been colour-coded to make these categories more visible.

Comparing the two negativity metrics provides a deeper understanding of how each topic affects negative relationships. For example, the words "peggio" and "gemist" are keywords in the topics most likely to be in a negative tweet, for the Italian and Dutch Generic Users datasets respectively (Table 18). This is unsurprising given that they mean "worse" or "the worst" in Italian and "excrement" in Dutch. However, they drop down to the 4^{th} and 6^{th} most negative topics in terms of impact on a user's relationships (Table 17). Revealing that, although they are frequently used in negative contexts, their influence in terms of relationships is lower than expected. By comparison, "salvini" and "biden", both well-known politicians, have the greatest negative impact on users' relationships out of any topics of the Italian and Dutch datasets while being at the 4^{th} and 6^{th} positions for tweet negativity. This shows that, although they are less frequently used in negative contexts than "peggio" and "gemist", they are much stronger indicators of negativity in a user's surrounding network.

Subsequently, the negativities of the top topics are viewed when grouped into the aforementioned categories (Politics, Religion, Football, COVID and Generic). As was just touched upon, Politics is by far the most negative category of topic in all three datasets. They are also the most specific topics, with all of the topics referencing specific people (Matteo Salvini, Matteo Renzi, Vladimir Putin and Joe Biden), places (The Hague) or problems (Operação Lava Jato and rising fuel prices/Petroleo Brasileiro SA). The only non-specific political topic is "democrazia" ("democracy" in Italian), which is also the least negative (in terms of placement). Thus, the higher negativity of the political

Table 17: Top 20 topics for the 3 Generic Users datasets and the mean percentage of negative relationships of users who are engaged with each topic, ordered by negativity. The topics are colour-coded: red for Politics, green for COVID, yellow for Religion, purple for Football and blue for Generic.

	Italy (66.75%)		Brazil (64.58%)		Netherlands (54.55%)	
Index	Topic	Negativity	Topic	Negativity	Topic	Negativity
1	salvini	89.17	lava	81.03	biden	64.03
2	renzi	86.88	putin	80.50	haag	61.42
3	amazon	79.07	gasolina	76.93	lachen	60.05
4	peggio	77.89	menino	76.49	nope	59.96
5	odio	76.97	verdades	75.93	hond	58.77
6	democrazia	76.93	jesus	75.53	gemist	57.07
7	nomi	75.03	cabelo	75.45	trein	56.42
8	pizza	74.75	perdi	74.88	vakantie	55.90
9	virus	74.26	fã	73.44	slapen	55.74
10	papa	74.17	festa	73.00	coronavirus	54.74
11	concordo	73.71	meme	72.77	filmpje	53.18
12	calcio	72.99	máscara	72.01	gold	51.53
13	thread	69.35	netflix	71.13	aflevering	51.15
14	dibattito	68.54	barato	70.99	facebook	50.93
15	caffè	66.71	gato	68.91	seizoen	50.38
16	natale	66.47	pizza	67.54	koffie	49.78
17	sogno	66.28	facebook	67.52	verjaardag	48.69
18	coronavirus	66.25	artista	66.11	interviews	48.40
19	facebook	63.84	natal	64.99	fotos	46.71
20	serata	62.33	dm	63.96	anniversary	40.22

Table 18: Top 20 topics for the 3 Generic Users datasets and the mean negativity of all corresponding tweets, ordered by negativity. The topics are colour-coded: red for Politics, green for COVID, yellow for Religion, purple for Football and blue for Generic.

	Italy (66.75%)		Brazil (64.58%)		Netherlands (54.55%)	
Index	Topic	Negativity	Topic	Negativity	Topic	Negativity
1	peggio	95.33	putin	85.59	gemist	55.59
2	odio	90.09	perdi	82.11	coronavirus	45.90
3	virus	83.03	lava	68.50	haag	45.15
4	salvini	81.18	jesus	59.87	hond	41.87
5	renzi	80.94	máscara	59.01	trein	37.23
6	democrazia	60.09	verdades	49.71	biden	36.02
7	coronavirus	59.40	gasolina	48.55	lachen	34.38
8	papa	54.53	menino	47.64	slapen	34.31
9	calcio	53.02	meme	41.45	nope	29.83
10	nomi	51.67	gato	40.71	vakantie	25.66
11	pizza	47.01	cabelo	37.08	facebook	25.46
12	amazon	43.00	festa	36.05	filmpje	23.18
13	concordo	38.53	fã	35.97	gold	19.07
14	dibattito	33.88	facebook	35.67	seizoen	16.92
15	caffè	32.23	barato	35.27	fotos	16.88
16	facebook	30.36	netflix	34.91	interviews	16.75
17	natale	28.42	pizza	32.03	koffie	16.13
18	thread	26.81	artista	30.95	aflevering	11.59
19	sogno	24.39	natal	22.58	anniversary	10.16
20	serata	11.38	dm	16.74	verjaardag	8.72

category of topics may be due to their specificity, which, as previously mentioned, can lead to a greater number of conflicting opinions.

Next, the two smallest categories: Religion and Football. Topics relating to these categories only appeared in the Italian and Brazilian datasets for the former and in the Italian for the latter. The religious topics were focused on specific individuals (the pope and Jesus) whereas Football is a generic topic relating to the sport as a whole ("calcio" being the Italian for football). Despite the religious terms being just as specific as the political ones, they are visibly less negative overall, suggesting that specificity alone is not enough to explain the differing negativities. Football appears just below the corresponding religious topic in its dataset in both of the tables.

The COVID category shows the biggest difference between the two tables. In Table 17 these topics appear towards the lower half of the table and, the least negative non-Generic term in each of the three datasets belongs to the COVID category. However, in Table 18, it is almost the exact opposite, with the most negative non-Generic term belonging to this category for both the Italian and Dutch datasets. This difference between the tables may suggest that, while individual Tweets related to COVID do tend to be very negative, tweeting negatively about COVID does not have an overly strong negative effect on an individual's surrounding relationships. This highlights the difference between a negative topic and a controversial one. Despite COVID being a very negative category, there appear to be fewer differing opinions about it (i.e. there is a fairly strong consensus that COVID overall is bad) and, therefore, fewer disagreements compared to the other selected categories.

Counting the total number of topics in the final category, Generic, for each dataset, Italy has the fewest (13), followed by Brazil (15) and then the Netherlands (17). These numbers are reversely correlated with the order of the overall negativities of these datasets, which are 66.75%, 64.58% and 54.55% respectively. Thus, it appears that the more specific topics a user is engaged with, the higher the percentage of negative relationships they have is expected to be. This further supports previous work, which came to the same conclusion (Tacchi et al., 2023).

Finally, the categories of topics were grouped together and their mean values were calculated in terms of both tweet negativity (Table 18) and user negativity (Table 17). These values are displayed in Table 19.

Similar to the previous results, Politics is the most negative overall and is the most negative for all columns except the Dutch Tweet negativity, where it is beaten by COVID. Religion is usually the second most negative for the two datasets in which it appears, Italy and Brazil, and it is followed by Football in the former. COVID is the least consistent, appearing in 4 of the 5 possible positions. Finally, Generic

Table 19: Mean user and tweet negativities, by category, of the top 20 topics in the 3 novel Generic Users datasets, ordered by negativity. The categories are colour-coded: red for Politics, green for COVID, yellow for Religion, purple for Football and blue for Generic.

Italy		Brazil		Netherlands	
User	Tweet	User	Tweet	User	Tweet
84.32	74.07	79.48	67.55	62.73	45.90
74.17	71.22	75.53	59.87	54.74	40.59
72.99	54.53	72.01	59.01	52.64	26.10
70.84	53.02	70.87	38.24	–	–
70.25	39.91	–	–	–	–

is the least negative, appearing in the least negative position for all columns except the Italian user negativity, where it appears above COVID.

As these are the means of other values previously discussed in detail, it is unsurprising that they do not provide any additional results. Their main value is to confirm the overall negativity of very controversial topics (such as Politics), and the lower negativity of neutral topics (Generic) with respect to the more specific topics (Football, Religion, COVID).

6.4 CHAPTER SUMMARY

This chapter has built upon the foundations laid by the previous structural investigations of the SENM by comparing differences across different cultures and online communities. This has resulted in observations that both a group's culture and the main topics its members engage with are important factors for the amount of negative relationships that are present. What's more, the more specific a topic is, the more negativity it seems to elicit, whereas engaging in more general topics, such as the weather, appears less likely to lead to negative relationships. Subsequently, a more targeted analysis of hashtags, words and topics revealed that these effects can be observed, even within subcommunities of generic users. Finally, comparisons of user and TwFeet negativities show that engaging with certain topics can have a negative influence on the sentiment of individual interactions, without a proportional negative effect on a person's relationships, and vice versa.

7 CONCLUSION

This chapter closes the thesis by providing an overview summary of its main results and contributions, followed by a discussion of some potential future avenues of research based on the topics discussed.

7.1 SUMMARY

The main results and contributions of this thesis are ordered by chapter. In chapter 2 an extensive collection of 11 curated preexisting datasets is presented. These datasets cover a variety of specialised and generic users and are further supplemented by 24 additional, novel datasets, collected specifically for the work within this thesis. Each of these latter datasets was iteratively selected to further explore previously observed phenomena. They represent a systematic cross-section of users from 4 geographically distinct cultures, all engaged in topics of varying levels of polarisation. The Tweet IDs of every post made by every user in this collection have been publicly shared, in compliance with X's terms of service. Therefore making it possible for anyone to view, recollect and reuse all of the data. Although, as previously mentioned, recent changes to the availability of the X API added significant obstacles to this.

Then, chapter 3 proposes a novel method of generating signed labels for entire relationships. This methodology was then applied to the aforementioned datasets and the resulting labels were compared using 4 different sentiment analysis models. The generated signs were remarkably consistent across different models, showing that the method achieves similar results regardless of the choice of model. These results were then validated against known expectations of signed networks. It was found that the distribution of the signs varied significantly from random, in the directions predicted by Balance Theory, lending a great deal of support to the validity of the generated signs and the signing method.

Chapter 4 then uses the newly generated signs to observe the SENMs within each of the datasets. First, the full and active networks were compared, whereby it was found that, surprisingly, the most negative relationships appear to be predominantly inside the active network. Similarly, when comparing levels of negativities across the different layers of the SENM, the inner circles were found to be the most negative. These 2 findings suggest that, at least on the X social media platform, the more time users spend engaging with a relationship, the more likely that relationship is to be negative. Otherwise said,

users interact most with negative connections. These findings were then followed up with an investigation of the interaction between negativity and cognitive load. Three different metrics were used to measure cognitive load. It was found that, although some minor effects may be present, the evidence was not significant enough to substantiate any claims of the relationship between negativity and cognitive load.

Next, chapter 5 investigated a promising application of the ENM and SENM: stance detection. This task was identified based on previous findings in adjacent network-related research areas that have demonstrated the potential benefits of using structural network features, such as the ENM. Although the results were slightly worse than the previous cutting-edge model, CT-TN, the ENMs only required data that is much easier to obtain. Thus, highlighting the potential of the ENM and SENM for similar applications to social network tasks, even where data is limited.

Finally, chapter 6 studied differences in SENM negativity between various combinations of culture and online communities. It was found that both culture and community membership had an effect on expected levels of negativity, with culture being the most impactful of the two. A pair of subsequent investigations of the most popular words and topics being discussed in some of the datasets shows that engaging more with polarising topics leads to more negative relationships. What's more, engaging with negative, although less controversial topics (i.e. topics that everyone agrees are bad, such as COVID-19) has a noticeably smaller impact on the rates of negative relationships, even though the Tweets related to such subjects may be far more negative than other topics.

7.2 FUTURE WORK

In addition to the concrete contributions detailed above, this thesis has also provided several promising directions for future work. Indeed, given that the proposed method of computing signs for relationships can be applied to any dataset that has text-based interactions between users, this method could easily be applied to many more tasks that did not fall within the scope of this thesis. For instance, as mentioned in Section 1.3, community detection, information diffusion and opinion dynamics are all tasks that can benefit from having the additional information provided by signed edges. A systematic exploration of such problems could also reveal which tasks benefit the most from signs and, perhaps, also why that is the case, furthering our understanding of the tacit information that signs provide.

Similarly, social networks from further sources could also be examined and compared to the current findings. Although, at the time of collection, X was by far the most open and popular social media

platform, researching other sources would provide a more complete idea of the phenomena observed. For example, as X is known for being a platform where negativity propagates far more easily than positivity Schöne, Parkinson, and Goldenberg, 2021, some of the negativity values may be slightly lower in different contexts. Despite this, given the strong theoretical foundation of anthropological and psychological research, the results are most likely still extremely generalisable.

Next, the cultural examinations in this thesis were conducted recursively, building upon the findings of previous investigations to inform each subsequent iteration. However, given the near-infinite depth and diversity of cultural differences, further investigations into differences and communications across different cultures may prove an endless source of further information. So, while the regions selected for analysis within this thesis represent a reasonable spread of different cultures, some regions of note were not able to be feasibly included. Unfortunately, while access to the internet is generally increasing and communication barriers are being slowly eroded, there will always be certain groups that provide much higher quantities of more easily accessible data than others, resulting in over- and under-represented groups. A clear example of this is India, where there was a relatively large community of X users at the time the data used in this thesis was collected. However, as XLM-T is known to particularly struggle with processing text in Hindi Barbieri, Anke, and Camacho-Collados, 2021, it was not possible to accurately include Indian data. Hopefully, this limitation can be born in mind during future research and, combined with improved language processing technologies, its impact can be reduced as much as possible.

Finally, as the proposed signing method was developed relatively recently, there is surely further room for refinement. For example, the 5:1 psychology-based threshold originally referred to negative interactions, however, negative sentiment was used as a proxy for this in the proposed signing methodology. With recent advances in Natural Language Processing (NLP), most notably Large Language Models (LLMs), which are able to deal with a broad range of language-related tasks, it could therefore be possible to directly ask the question: "Is this a negative interaction?".

APPENDIX A

A.1 LIST OF PUBLICATIONS

Here, all the publications that resulted from the work conducted as part of these thesis are listed, along with their relevant metadata (authors, date, etc.) and, where applicable, the state of their acceptance at time of writing.

Journal Papers

1. **Jack Tacchi**, Chiara Boldrini, Andrea Passarella, Marco Conti. "Keep Your Friends Close, and Your Enemies Closer: Structural Properties of Negative Relationships on Twitter." ACCEPTED AT: EPJ Data Science, April 2024.
2. **Jack Tacchi**, Chiara Boldrini, Andrea Passarella, Marco Conti. "On the Joint Effect of Culture and Discussion Topics on X (Twitter) Signed Ego Networks." SUBMITTED TO: PLOS One, June 2024.

Conference Papers

1. **Jack Tacchi**, Chiara Boldrini, Andrea Passarella, Marco Conti. "Signed ego network model and its application to Twitter." 2022 IEEE International Conference on Big Data (Big Data). IEEE, 2022.
2. **Jack Tacchi**, Chiara Boldrini, Andrea Passarella, Marco Conti. "Cultural Differences in Signed Ego Networks on Twitter: An Investigatory Analysis." Companion Proceedings of the ACM Web Conference 2023. 2023.
3. **Jack Tacchi**, Parisa Jamadi Khiabani, Arkaitz Zubiaga, Chiara Boldrini, and Andrea Passarella. "Applying the Ego Network Model to Cross-Target Stance Detection." ACCEPTED AT: International Conference on Advances in Social Networks Analysis and Mining, 2024.

APPENDIX B

B.1 FULL AND ACTIVE NETWORK NEGATIVITIES

The mean percentages of negative relationships in the full and active networks, as well as the difference in percentage points between them. Unfortunately, given the time required to perform sentiment analysis, it was not possible to obtain the sentiments for the full networks of the Politics, Football, Weather and Country Generic datasets (discussed in Section 4.1). However, the percentages were obtained for active networks of all datasets used in this thesis.

Table 20: Mean percentages of negative relationships in the full and active networks. A dashed line is used to separate specialised users (above) and generic users (below).

Dataset	Full Negatives (%)[a]	Active Negatives (%)[a]	Difference
American Journalists	27.15 [27.06, 27.23]	47.97 [47.71, 48.23]	+20.82
Australian Journalists	31.54 [31.41, 31.67]	54.39 [54.03, 54.75]	+22.85
British Journalists	28.78 [28.62, 28.94]	50.37 [49.89, 50.85]	+21.59
Italian Journalists	40.61 [40.45, 40.77]	63.87 [63.43, 64.31]	+23.26
Brazilian Journalists	44.80 [44.67, 44.92]	64.93 [64.49, 63.37]	+20.13
Dutch Journalists	34.55 [34.50, 34.60]	57.65 [57.47, 57.83]	+23.10
NYT Journalists	31.58 [31.43, 31.74]	54.89 [54.49, 55.29]	+23.31
Science Writers	25.62 [25.45, 25.78]	45.23 [44.71, 45.75]	+19.62
Monday Motivation	16.45 [16.36, 16.54]	21.83 [21.58, 22.08]	+5.38
UK Users	24.22 [24.13, 24.31]	35.32 [35.04, 35.60]	+11.10
British MPs	19.24 [19.09, 19.38]	29.03 [28.66, 29.39]	+9.79
Baseline	24.05 [24.03, 24.08]	40.31 [40.19, 40.44]	+16.26
Mediterranean	45.76 [45.69, 45.83]	60.08 [59.85, 60.31]	+14.32
South America	42.83 [42.72, 42.93]	65.67 [65.33, 66.01]	+22.85
Northern Europe	38.31 [38.23, 38.40]	54.66 [54.37, 54.97]	+16.34
West Africa	32.77 [32.69, 32.85]	50.29 [49.98, 50.60]	+17.52
Italian Reality TV	52.13 [52.01, 52.24]	64.97 [64.48, 65.48]	+12.84
Brazilian Reality TV	44.18 [44.07, 44.29]	69.47 [69.03, 69.91]	+25.29
Dutch Reality TV	46.39 [46.34, 46.45]	68.36 [68.18, 68.52]	+21.97
Italian Politics	-	84.34 [84.19, 84.49]	-
Brazilian Politics	-	81.24 [80.88, 81.60]	-
Dutch Politics	-	78.25 [78.04, 78.45]	-
Nigerian Politics	-	68.30 [67.95, 68.66.]	-
Italian Football	-	77.96 [77.74, 78.18]	-
Brazilian Football	-	71.41 [71.15, 71.67]	-
Dutch Football	-	64.09 [63.85, 64.33]	-
Nigerian Football	-	58.45 [57.64, 59.26]	-
Italian Weather	-	74.24 [73.82, 74.66]	-
Brazilian Weather	-	69.88 [69.45, 70.32]	-
Dutch Weather	-	56.67 [55.86, 57.48]	-
Nigerian Weather	-	60.73 [60.09, 61.37]	-
Italian Generic	-	66.75 [66.57, 66.93]	-
Brazilian Generic	-	64.58 [64.48, 64.69]	-
Dutch Generic	-	54.55 [54.45, 54.66]	-

[a]95% confidence intervals shown in square brackets.

APPENDIX C

C.1 NEGATIVITY METRICS BOXPLOTS

Boxplots displaying the 3 negativity metrics compared against the mean active Ego Network size (left column) and the mean number of interactions (right column), binned into quantiles. This analysis was conducted, in Section 4.3, for each of the monolingual English datasets. The boxplots display mean (orange line), median (green triangle), first to third quartile (box), 1.5 times the interquartile range beyond the box (whiskers) and outliers (black circles).

For the benefit of the reader, the definitions of the 3 metrics are repeated here:

1. The number of negative relationships that each Ego had, divided by their total number of relationships
2. the number of negative interactions for each Ego, divided by their total number of interactions
3. the number of each Ego's interactions that correspond to a negative relationship, divided by their total number of interactions

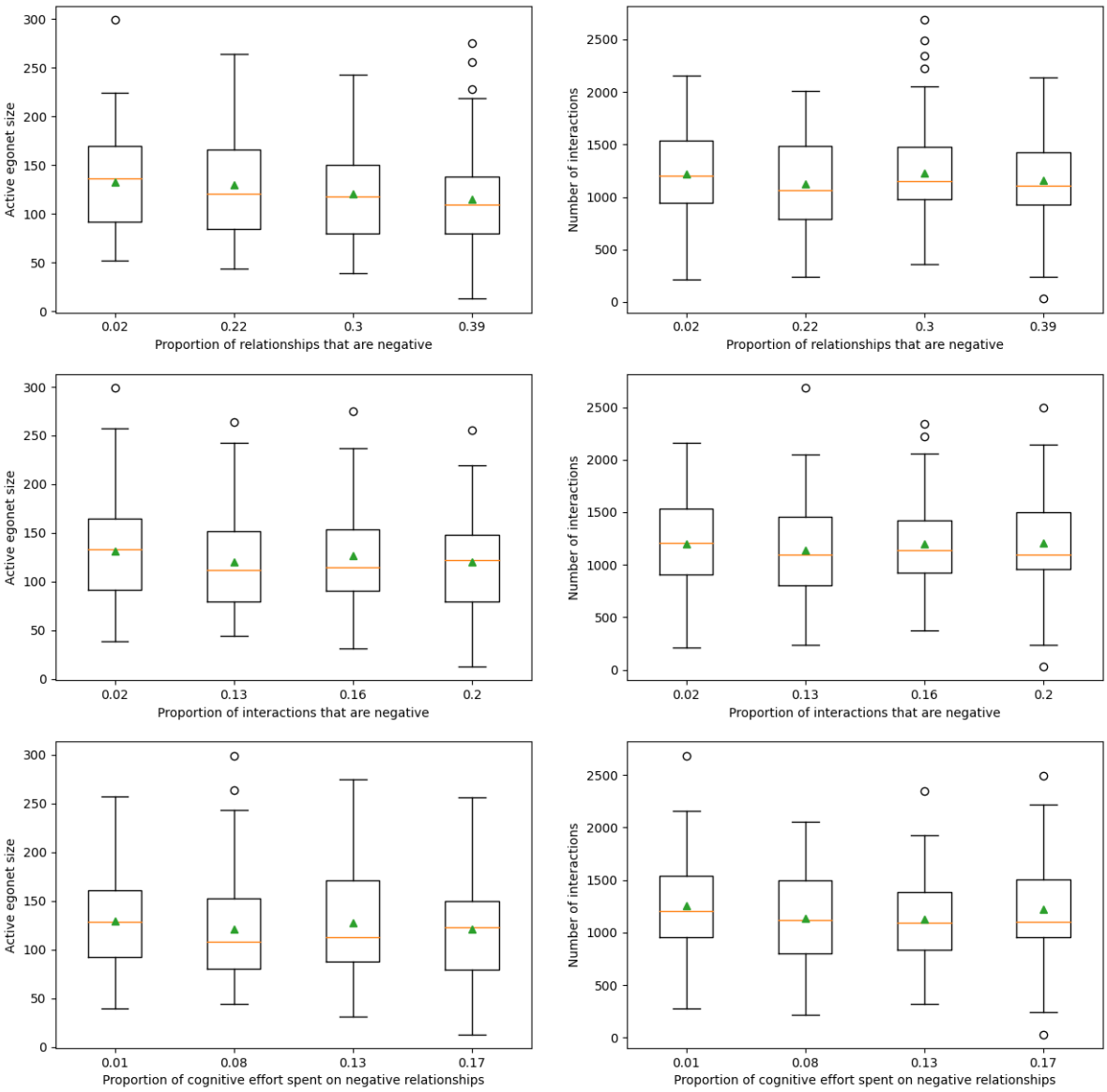

Figure 18: Boxplots for the American Journalists dataset.

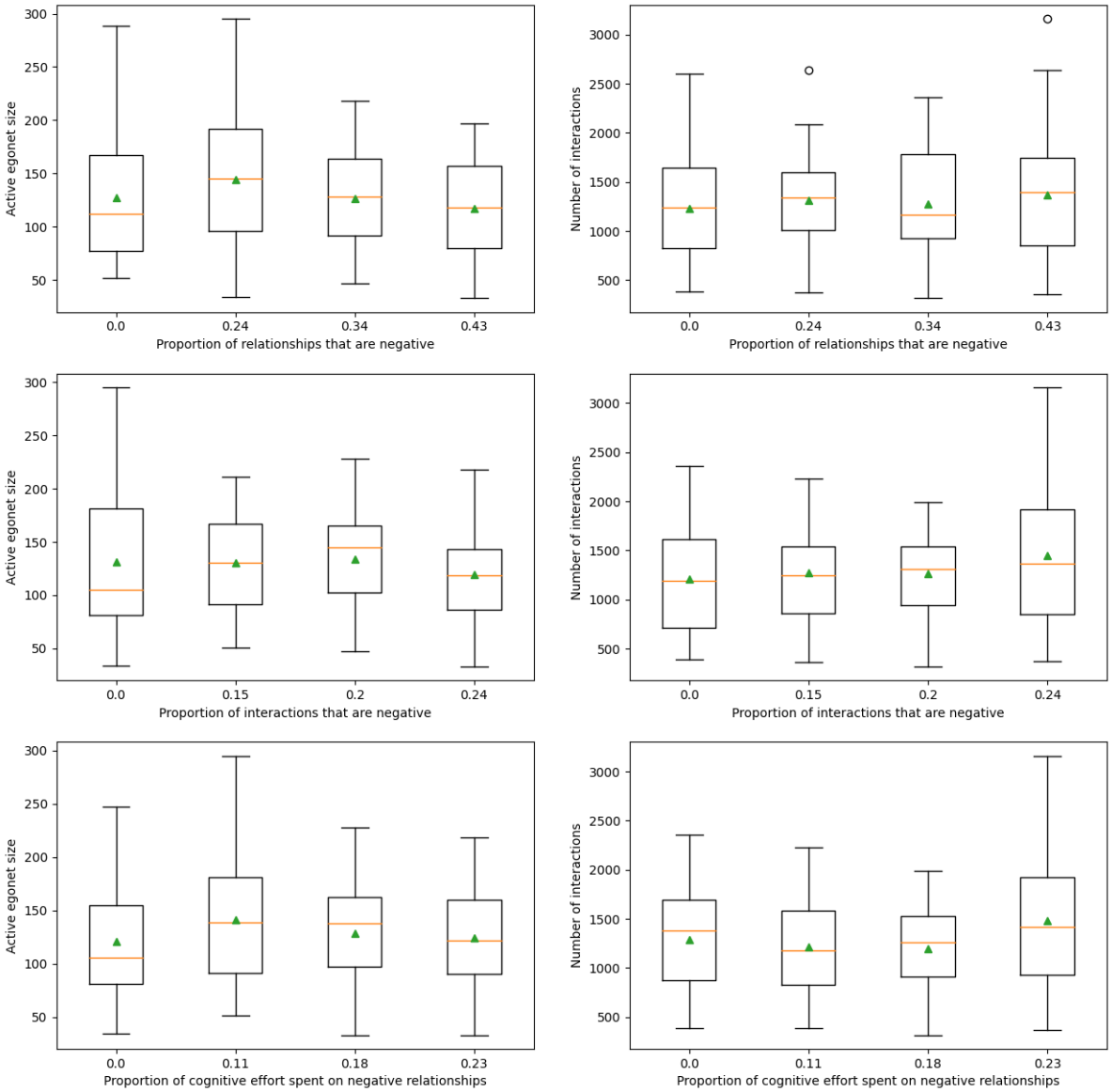

Figure 19: Boxplots for the Australian Journalists dataset.

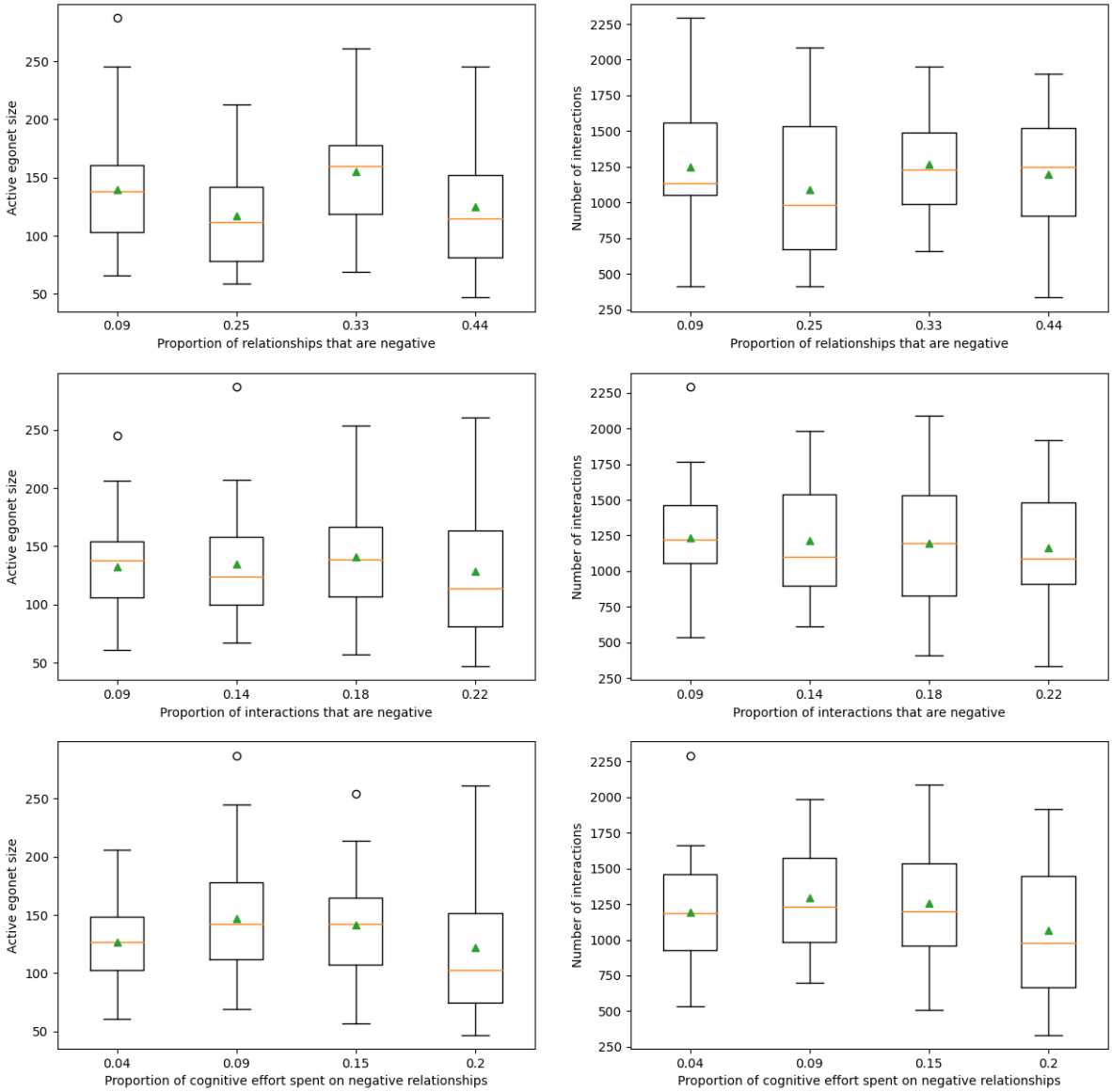

Figure 20: Boxplots for the British Journalists dataset.

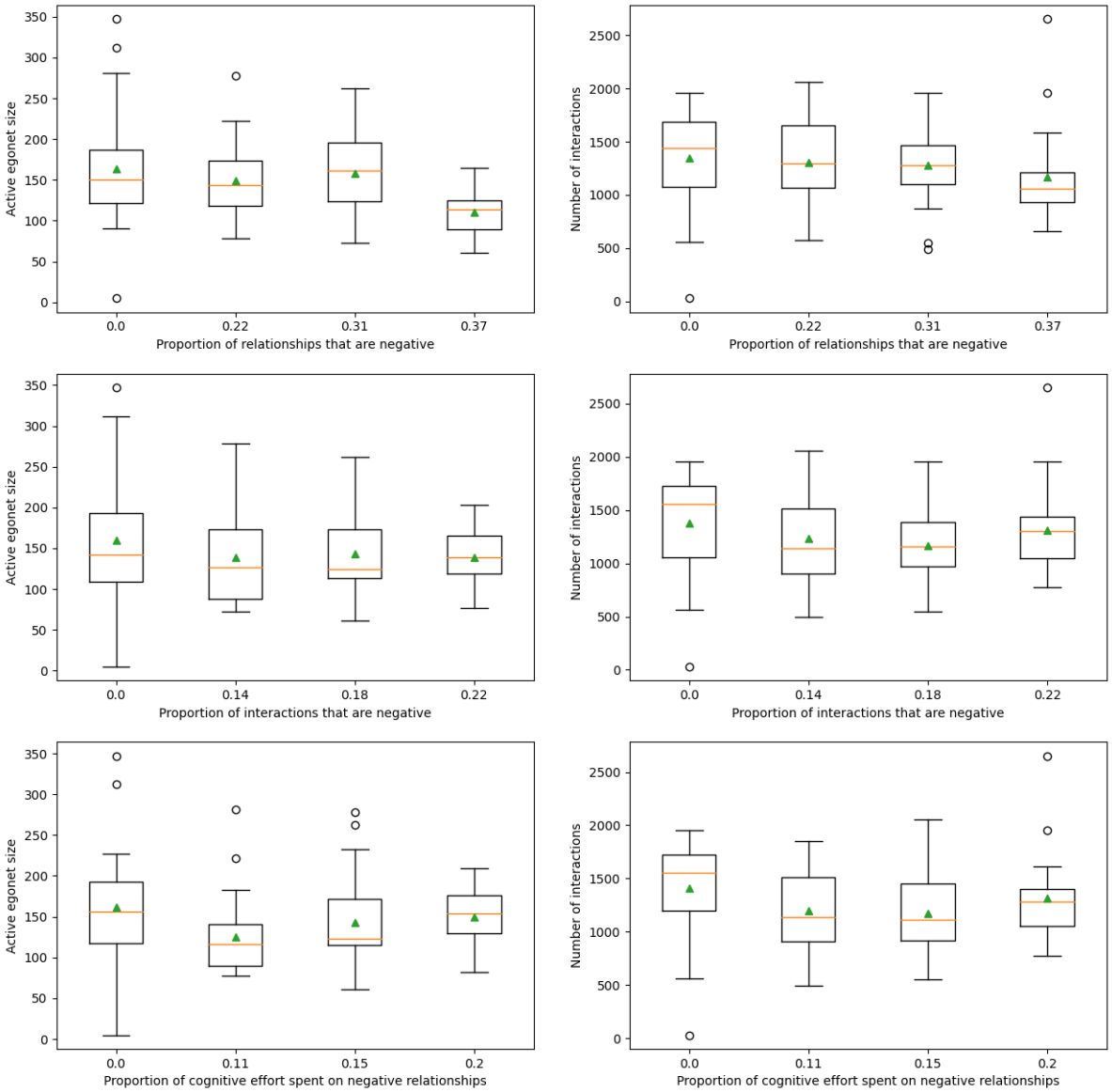

Figure 21: Boxplots for the NYT Journalists dataset.

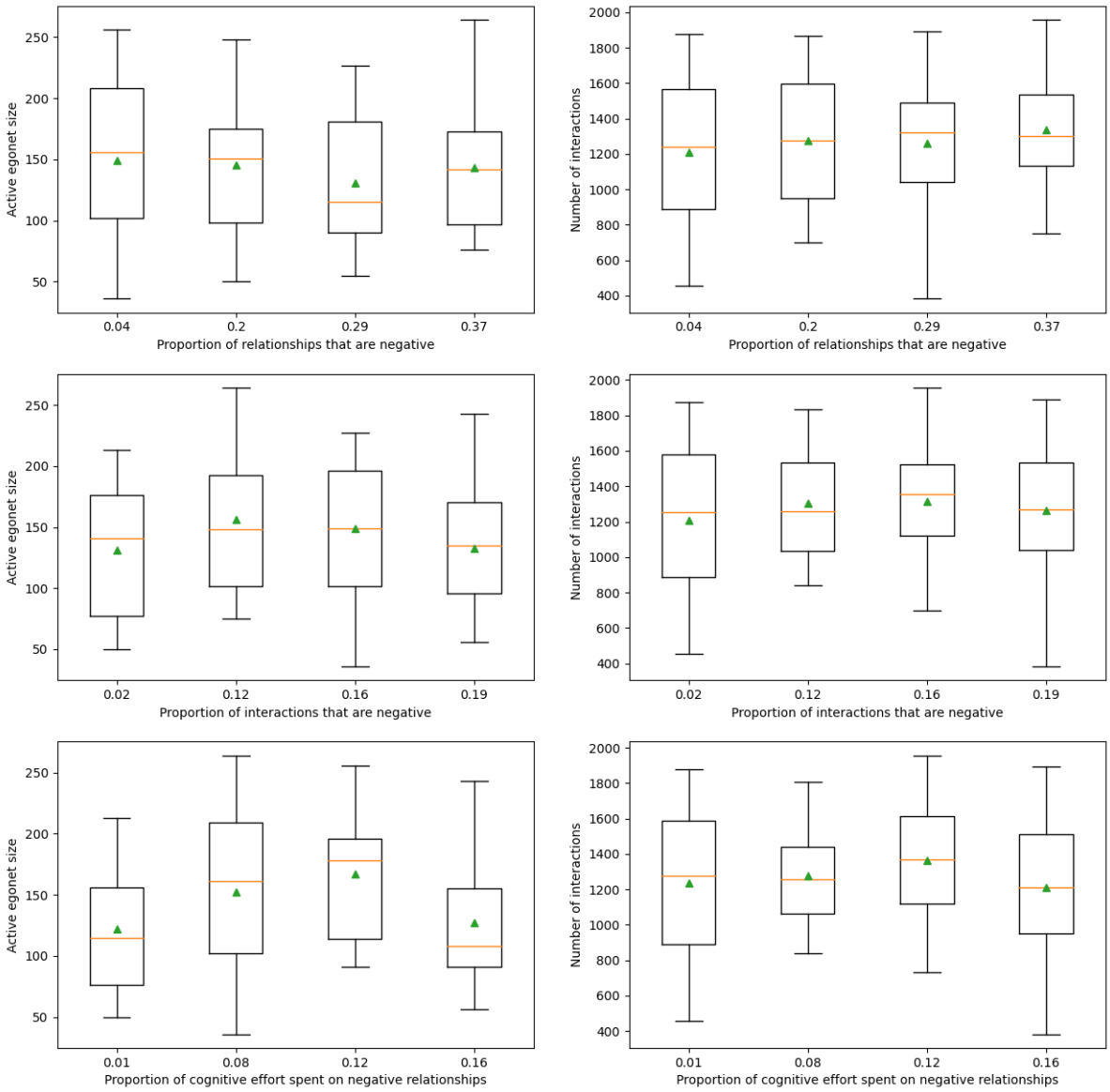

Figure 22: Boxplots for the Science Writers dataset.

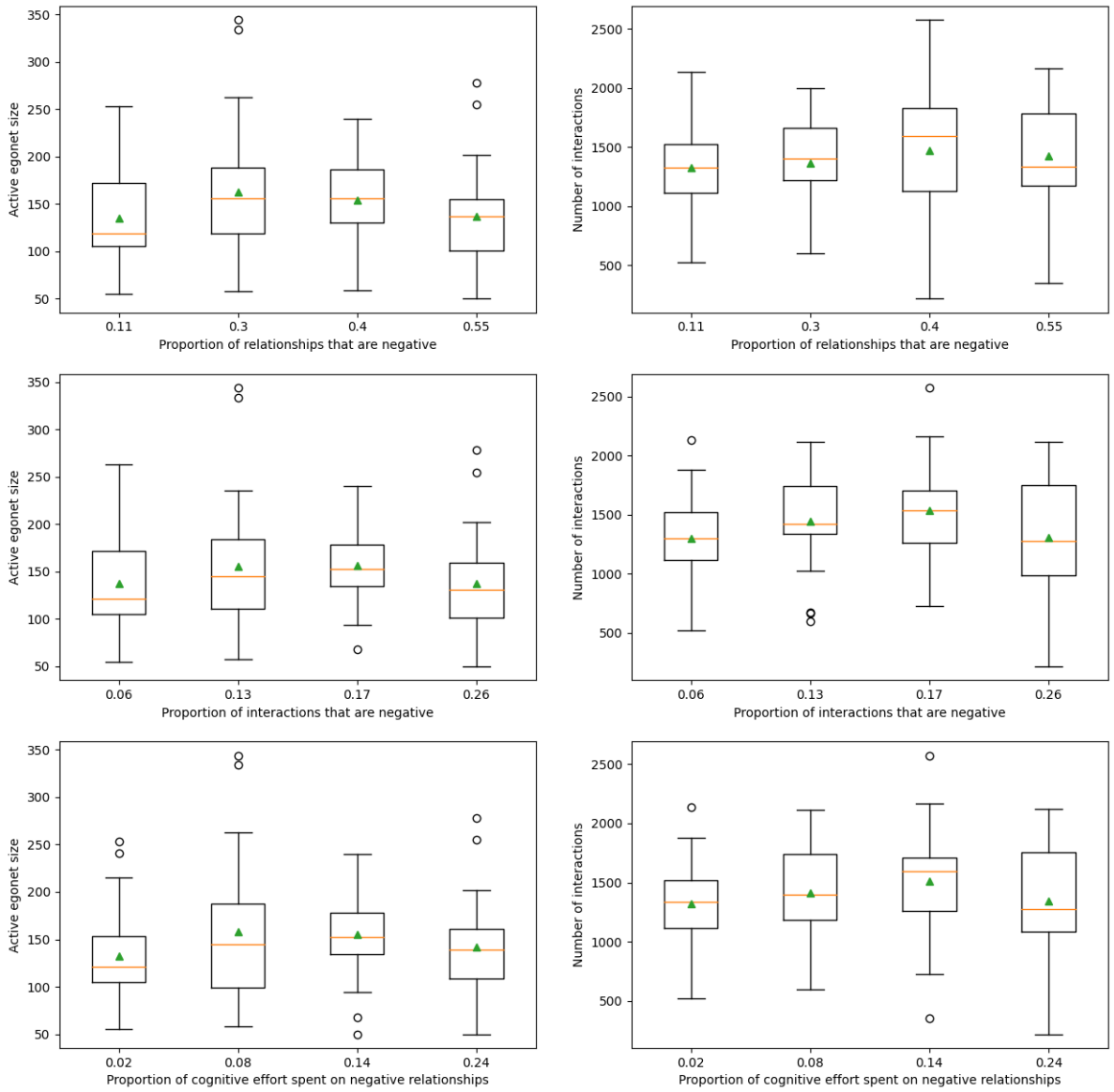

Figure 23: Boxplots for the British MPs dataset.

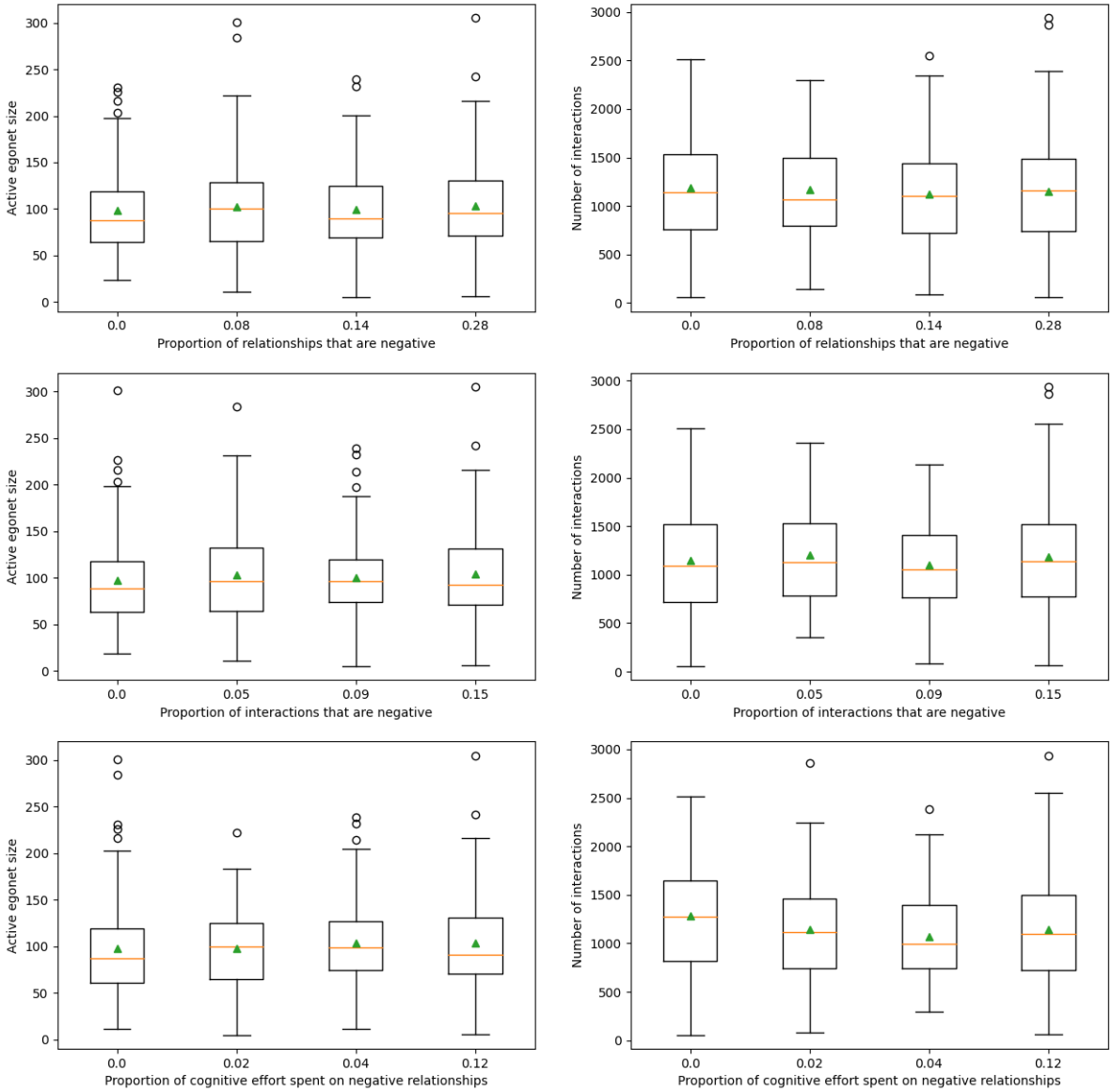

Figure 24: Boxplots for the Monday Motivation dataset.

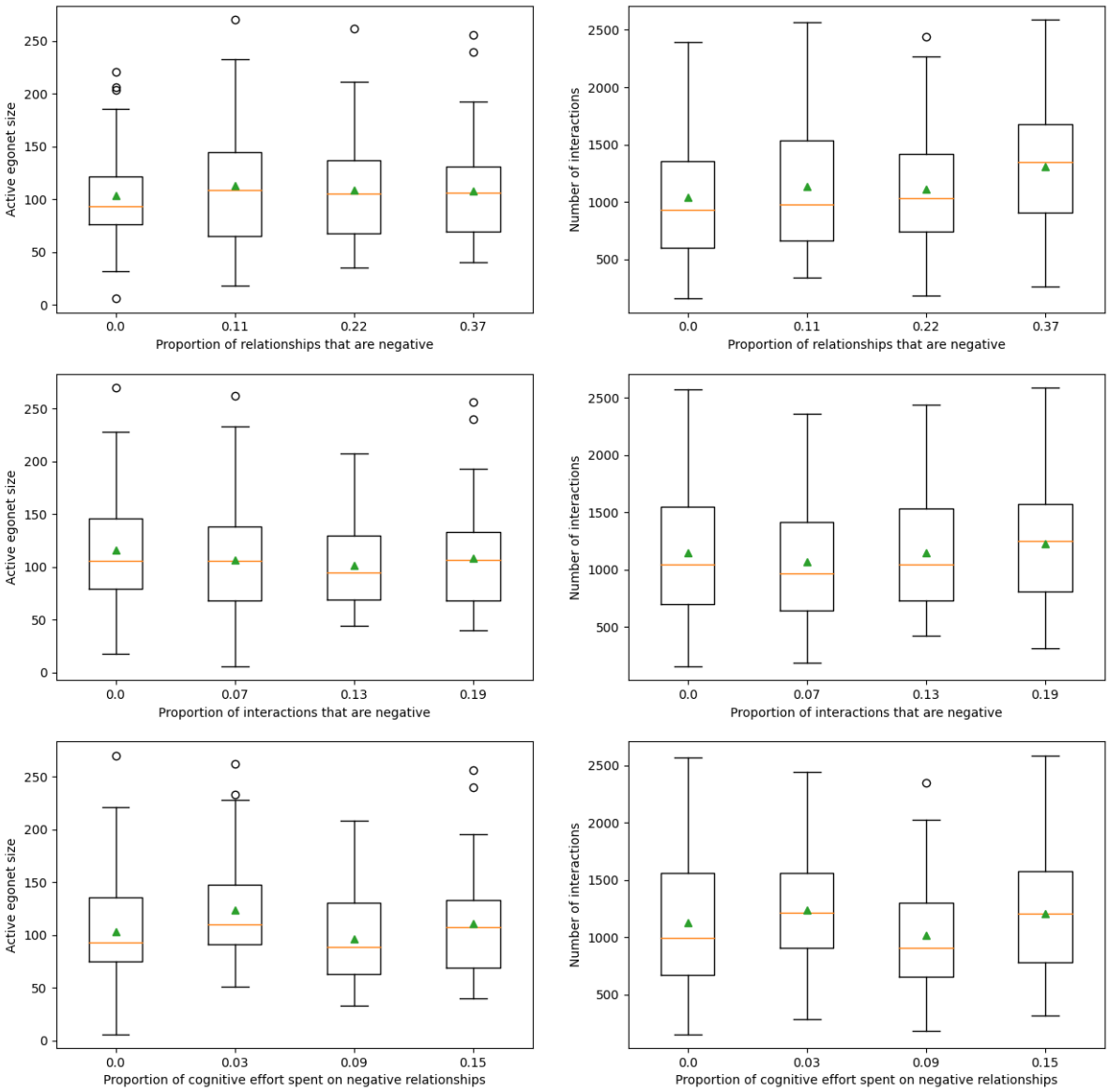

Figure 25: Boxplots for the UK Users dataset.

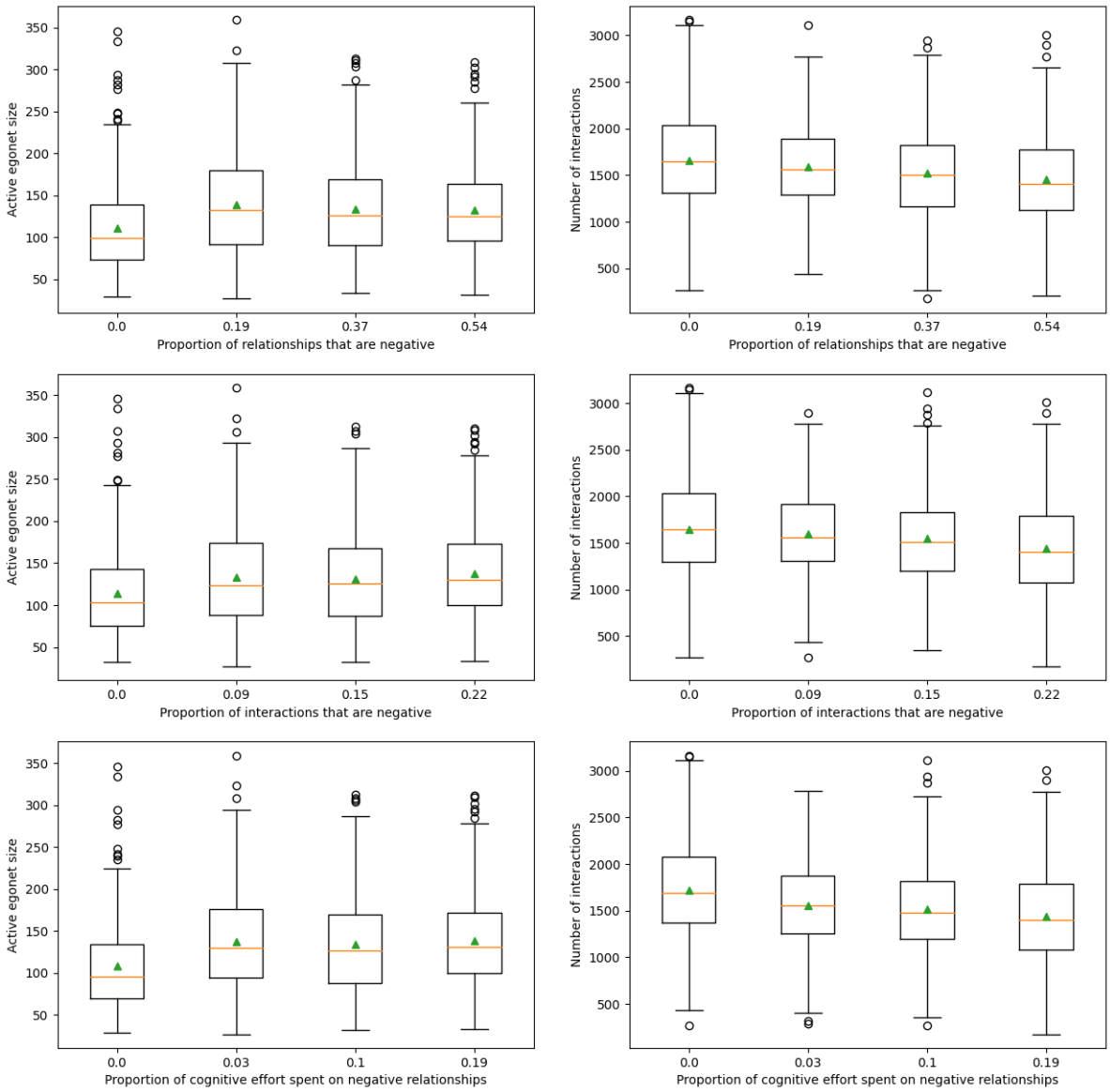

Figure 26: Boxplots for the Baseline dataset.

APPENDIX D

D.1 NEGATIVITY METRIC T-SCORES

The t-scores of the negativity metric t-tests (conducted in Section 4.3). Table 21 displays the results for the 3 metrics compared against the sizes of users' Ego Networks and Table 22 shows the same but compared against the number of users' interactions.

For the benefit of the reader, the definitions of the 3 metrics are repeated here:

1. The number of negative relationships that each Ego had, divided by their total number of relationships
2. the number of negative interactions for each Ego, divided by their total number of interactions
3. the number of each Ego's interactions that correspond to a negative relationship, divided by their total number of interactions

Table 21: The t-scores from the pairwise comparisons between bins for Ego network sizes and negativity. Values corresponding to statistically significant p-values are displayed in bold.

		Bin pairs					
	Dataset	**1-2**	**1-3**	**1-4**	**2-3**	**2-4**	**3-4**
Metric 1	American Journalists	0.328	1.440	**2.198**	1.053	1.790	0.790
	Australian Journalists	-1.197	0.043	0.822	1.480	**2.306**	0.974
	British Journalists	1.426	-0.938	0.886	**-2.503**	-0.468	1.845
	NYT Journalists	0.794	0.302	**3.361**	-0.606	**3.479**	**4.045**
	Science Writers	0.191	0.884	0.291	0.770	0.111	-0.660
	British MPs	-1.553	-1.332	-0.121	0.514	1.459	1.214
	Monday Motivation	-0.529	-0.136	-0.762	0.413	-0.195	-0.646
	UK Users	-0.997	-0.602	-0.560	0.429	0.462	0.037
	Baseline	**-5.657**	**-4.595**	**-4.471**	1.226	1.456	0.216
Metric 2	American Journalists	1.205	0.552	1.397	-0.656	0.105	0.808
	Australian Journalists	0.070	-0.222	0.875	-0.370	1.012	1.439
	British Journalists	-0.162	-0.552	0.247	-0.350	0.360	0.688
	NYT Journalists	1.072	0.897	1.200	-0.271	-0.042	0.294
	Science Writers	-1.193	-0.885	-0.051	0.373	1.219	0.894
	British MPs	-0.962	-1.323	0.029	-0.077	1.004	1.390
	Monday Motivation	-0.866	-0.478	-1.010	0.443	-0.150	-0.599
	UK Users	0.987	1.714	0.869	0.588	-0.158	-0.799
	Baseline	**-3.926**	**-3.593**	**-4.926**	0.431	-0.775	-1.249
Metric 3	American Journalists	0.992	0.266	1.098	-0.659	0.006	0.719
	Australian Journalists	-1.481	-0.675	-0.280	0.968	1.366	0.457
	British Journalists	-1.363	-1.061	0.264	0.324	1.305	1.054
	NYT Journalists	**2.061**	1.007	0.771	-1.134	-1.977	-0.476
	Science Writers	-1.506	**-2.459**	-0.297	-0.710	1.288	**2.257**
	British MPs	-1.396	-1.669	-0.617	0.132	0.892	1.002
	Monday Motivation	0.004	-0.910	-0.862	-1.083	-1.007	0.003
	UK Users	**-2.127**	0.755	-0.799	**3.111**	1.363	-1.655
	Baseline	**-6.271**	**-5.463**	**-6.524**	0.805	-0.028	-0.859

Table 22: The t-scores from the pairwise comparisons between bins for the number of interactions and negativity. Values corresponding to statistically significant p-values are displayed in bold.

		Bin pairs					
	Dataset	**1-2**	**1-3**	**1-4**	**2-3**	**2-4**	**3-4**
Metric 1	American Journalists	1.362	-0.113	0.819	-1.453	-0.584	0.920
	Australian Journalists	-0.665	-0.380	-0.993	0.290	-0.444	-0.680
	British Journalists	1.149	-0.124	0.399	-1.304	-0.762	0.536
	NYT Journalists	0.384	0.593	1.385	0.229	1.112	0.902
	Science Writers	-0.446	-0.363	-1.014	0.089	-0.505	-0.612
	British MPs	-0.352	-1.192	-0.876	-0.899	-0.558	0.363
	Monday Motivation	0.241	0.967	0.509	0.746	0.279	-0.456
	UK Users	-0.908	-0.673	**-2.563**	0.292	-1.664	**-2.062**
	Baseline	1.530	**2.973**	**4.364**	1.743	**3.369**	1.602
Metric 2	American Journalists	0.726	-0.009	-0.112	-0.751	-0.842	-0.105
	Australian Journalists	-0.575	-0.518	-1.665	0.090	-1.230	-1.341
	British Journalists	0.161	0.277	0.518	0.136	0.369	0.199
	NYT Journalists	1.113	1.751	0.533	0.636	-0.695	-1.417
	Science Writers	-0.712	-0.780	-0.391	-0.087	0.329	0.404
	British MPs	-1.405	**-2.229**	-0.120	-0.769	1.077	1.786
	Monday Motivation	-0.669	0.715	-0.429	1.532	0.226	-1.214
	UK Users	0.738	0.009	-0.732	-0.761	-1.529	-0.776
	Baseline	1.103	**2.298**	**4.382**	1.386	**3.774**	**2.448**
Metric 3	American Journalists	1.592	1.713	0.440	0.063	-1.110	-1.210
	Australian Journalists	0.599	0.789	-1.355	0.168	-1.886	**-2.082**
	British Journalists	-0.811	-0.467	0.946	0.281	1.696	1.322
	NYT Journalists	1.729	1.819	0.776	0.174	-1.116	-1.240
	Science Writers	-0.308	-0.891	0.184	-0.779	0.576	1.198
	British MPs	-0.908	-1.683	-0.247	-0.784	0.580	1.310
	Monday Motivation	1.759	**2.978**	1.803	1.176	0.065	-1.092
	UK Users	-1.077	1.071	-0.726	**2.342**	0.372	-1.945
	Baseline	**4.007**	**4.700**	**6.148**	0.764	**2.595**	1.887

APPENDIX E

E.1 MOST POPULAR HASHTAGS AND WORDS

The top 20 most used hashtags and words of the active networks of the regional generic (including Baseline), non-English language journalists and Reality TV datasets (from Section 6.2). They are colour-coded by topic: "Political" in orange, "General" in blue, "COVID" in red, "Climate" in green, "Religious" in pink and "News" in grey. The percentages of negative relationships in the active networks of each dataset are also given alongside the names.

	Mediterranean (60.08%)		South America (65.67%)		Northern Europe (54.66%)		West Africa (50.29%)		Baseline (40.31%)	
Index	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts
1	#covid19	6,635	#afp	3034	#energiewende	2,519	#kebetu	2,792	#covid19	6,508
2	#bordeaux	2,127	#ultimahora	1398	#klimaschutz	2,196	#covid19	2,669	#np	6,370
3	#dimartedi	2,016	#envivo	1295	#berlin	1,886	#senegal	2,242	#ff	4,003
4	#coronavirus	1,915	#venezuela	1139	#corona	1,526	#voyagesafriq	1,866	#nowplaying	3,250
5	#qag	1,787	#legendarios	774	#svpol	1,486	#endsars	1,151	#quote	3,210
6	#...	1,686	#espnfcolombia	616	#nrw	1,467	#thegrill	1,150	#hiphop	2,934
7	#paris	1,566	#ahora	465	#solar	1,437	#weareicgc	886	#followfriday	2,930
8	#εμπιστευτικα	1,482	#cispaldia	461	#ukraine	1,253	#endinsecurity	861	#...	2,882
9	#directsenat	1,222	#espaciopolitico	454	#btw21	953	#music	747	#iem	2,658
10	#draghi	1,177	#envideo	352	#klimakrise	926	#putyoungpeoplefirst	706	#1	2,491
11	#nouvelleaquitaine	1,171	#colombia	325	#covid19	902	#digitalnigeria	691	#travel	2,378
12	#conte	1,161	#cubaviveytrabaja	312	#pv	882	#sundaysaticgc	644	#tni	2,262
13	#m5s	1,103	#esnoticia	311	#otd	867	#...	629	#tezos	2,170
14	#directan	1,093	#cubacoopera	304	#energytransition	848	#touchepasamasoeur	616	#win	2,107
15	#salvini	1,083	#enterate	295	#eu	819	#citizenlegs	563	#wwenxt	2,096
16	#lemissionpolitique	1,076	#fvmenlalucha	283	#betd2020	774	#nowplayingonwavefm	562	#giveaway	1,964
17	#governo	1,037	#contigo	269	#...	768	#radiostation	514	#nascar	1,917
18	#fonctionpublique	1,023	#video	267	#wnl	768	#endfgm	510	#business	1,867
19	#occitanie	1,003	#reportecovid19	264	#coronavirus	683	#generationequality	501	#caafb	1,821
20	#roma	952	#viacrucisdelmaestro	254	#renewables	683	#lunchtimevibes	495	#marketing	1,814

	Italian Journalists (63.87%)		Brazilian Journalists (64.93%)		Dutch Journalists (57.65%)	
Index	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts
1	#buongiorno	943	#brazil	663	#wnl	3,356
2	#agorarai	921	#marcocivil	651	#vught	1,584
3	#sanremo2016	726	#brazils	285	#rotterdam	1,450
4	#eniday	694	#politica	183	#eenvandaag	1,303
5	#torino	671	#brasil	177	#alphen	1,032
6	#primapagina	630	#internet	175	#ob	1,012
7	#trump	625	#tecnologia	160	#fd	993
8	#lintervista	623	#foratemer	154	#dtv	640
9	#usa2016	560	#j10	147	#vkopinie	627
10	#tg24pomeriggio	534	#arquivobbc	136	#china	605
11	#edicola	477	#worldcup	112	#ndnl	555
12	#quartogrado	475	#cartacapital	98	#bnr	524
13	#roma	456	#empauta	94	#nrc	519
14	#gruppoespresso	412	#dilma	85	#bd	464
15	#rai3	381	#redessociais	85	#mojo	458
16	#brexit	352	#g1	82	#brexit	457
17	#canale50	309	#ciudadesdemocraticas	79	#denhaag	450
18	#presson	297	#jornalismo	74	#nieuws	437
19	#milano	269	#conexoesglobais	73	#	376
20	#amorimoderni	268	#noticias	72	#schiphol	369

	Italian Reality TV (64.97%)		Brazilian Reality TV (69.47%)		Dutch Reality TV (68.36%)	
Index	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts
1	#prelemi	1,911	#xfactorbr	1,226	#gopcorruptionovercountry	445
2	#jeru	1,534	#bbb21	642	#ajax	444
3	#gfvip	1,093	#bbb22	293	#supplychain	416
4	#gfvipparty	736	#kep1er	261	#f1	377
5	#venice	569	#kepeulreo	237	#gopliesabouteverything	249
6	#venezia	490	#conceptacousticsessions	216	#gopbetrayedamerica	243
7	#taleequaleshow	440	#transbordoubrahmosidade	163	#voteblue2022	240
8	#ballandoconlestelle	435	#bts	161	#nde	218
9	#federicoangelucci	355	#gfriend	153	#groningen	207
10	#annalisa	343	#yeojacingu	137	#wweraw	194
11	#venise	342	#1	121	#smackdown	174
12	#venessia	334	#jimin	121	#sap	163
13	#venedig	333	#vmas	115	#bde	163
14	#venecia	327	#bangtansonyeondan	113	#gophypocrisy	155
15	#xf2020	322	#grammys	99	#neardeathexperience	143
16	#bellissima	285	#diaznoshoptime	96	#trumpcrimesyndicate	136
17	#chicagopd	284	#tokyo2020	85	#trumpisguilty	132
18	#...	278	#masterchefbr	77	#ruttemoetweg	121
19	#sanremo2021	271	#spacedomuka	69	#miasanmia	119
20	#mikainstagram	271	#wadada	68	#fcbayern	110

Figure 27: The 20 most used hashtags in the active networks.

	Mediterranean (60.08%)		South America (65.67%)		Northern Europe (54.66%)		West Africa (50.29%)		Baseline (40.31%)	
Index	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts
1	france	13,422	nicolasmaduro	14,411	heute	13,702	people	7,647	today	58,024
2	cette	12,862	venezuela	7,921	berlin	7,656	nigeria	5,949	people	51,285
3	c'est	11,614	presidente	4,893	schon	6,205	jeunes	5,694	great	49,611
4	merci	11,072	pueblo	4,573	annieloof	5,868	thank	5,645	thank	44,460
5	covid19	10,698	afpespanol	3,455	immer	5,433	today	5,621	first	43,433
6	emmanuelmacron	9,412	colombia	2,607	deutschland	5,405	covid19	5,461	would	42,222
7	faire	8,871	nacional	2,408	menschen	5,286	state	5,203	happy	37,390
8	essere	8,078	artecultor	2,217	müssen	5,159	femmes	4,910	thanks	37,325
9	sempre	7,933	gobierno	2,092	klimaschutz	4,533	women	4,506	think	33,484
10	governo	7,847	jguaido	2,046	energy	4,321	unfpawcaro	4,151	right	31,483
11	grande	7,734	maduro	2,035	morgen	4,029	please	4,070	tonight	31,478
12	stato	7,642	patria	2,006	viele	3,897	mbuhari	3,879	going	31,221
13	contre	7,521	dcabellor	1,997	tagesspiegel	3,871	president	3,811	still	30,585
14	είναι	7,521	henriolliveira	1,838	welt	3,779	allah	3,786	years	28,949
15	comme	7,386	mundo	1,777	danke	3,765	kindly	3,509	watch	27,102
16	fatto	7,370	junto	1,749	energiewende	3,765	happy	3,432	night	25,326
17	italia	6,972	jairbolsonaro	1,730	centerpartiet	3,300	sénégal	3,114	don't	25,226
18	président	6,971	gracias	1,729	corona	3,230	africa	3,070	really	25,017
19	https…	6,963	medicalara	1,685	jahren	3,226	world	3,054	never	23,282
20	senza	6,681	vamos	1,631	politik	3,203	first	2,843	could	22,461

	Italian Journalists (63.87%)		Brazilian Journalists (64.93%)		Dutch Journalists (57.65%)	
Index	Words	Counts	Words	Counts	Words	Counts
1	lastampa	23,831	stories	2,232	nieuwe	11,112
2	ilmessaggeroit	3,335	daily	1,652	vandaag	11,010
3	grazie	3,270	barca	1,640	volkskrant	10,819
4	domani	3,030	brasil	1,359	morgen	7,248
5	buongiorno	2,883	taborda	1,264	mensen	7,131
6	trump	2,734	brazil	1,208	alleen	6,727
7	nuovo	2,669	globonews	1,064	gewoon	6,526
8	prima	2,256	internet	1,023	jullie	6,514
9	corriere	2,247	temer	1,015	telegraaf	6,450
10	italia	2,196	reuters	891	staat	6,359
11	grande	1,986	agora	857	vanavond	5,968
12	skysport	1,918	dilma	767	waarom	5,646
13	sempre	1,896	latest	709	eenvandaag	5,479
14	renzi	1,876	governo	707	nederland	5,330
15	video	1,874	today	700	maken	5,277
16	lavoro	1,845	thanks	692	eerste	5,258
17	vaticanit	1,800	marcocivil	685	mooie	5,030
18	mondo	1,651	camarotti	621	moeten	5,027
19	senza	1,641	mundo	551	nieuws	5,022
20	raitre	1,608	radioesportesfm	541	zeker	4,986

	Italian Reality TV (64.97%)		Brazilian Reality TV (69.47%)		Dutch Reality TV (68.36%)	
Index	Hashtags	Counts	Hashtags	Counts	Hashtags	Counts
1	mikasounds	3,779	vídeo	5,308	mensen	2,519
2	nuovo	2,214	youtube	5,265	gewoon	2,274
3	video	2,207	gostei	5,022	everyanglesw	2,144
4	chiaragaliazzo	2,131	conceptofficial	1,694	alleen	1,729
5	madonna	2,099	arthurpicoli	1,358	lekker	1,604
6	francescacheeks	2,080	xfactorbr	1,314	jullie	1,497
7	fedez	2,074	gente	1,048	moeten	1,216
8	grazie	1,994	luablanco	928	hahaha	1,148
9	prelemi	1,947	xfactorbr	900	telegraaf	1,138
10	valerioscanu	1,945	btstwt	824	nunl	1,122
11	naliofficial	1,896	heyisis	801	maken	1,089
12	canzoni	1,859	gabzuski	776	nieuwe	1,056
13	sempre	1,806	arthur	728	nederland	1,041
14	https…	1,751	naiarazevedo	710	nooit	1,012
15	niallofficial	1,701	cover	669	helemaal	1,005
16	thisismaneskin	1,662	camilacabello	661	zeker	1,001
17	preferite	1,636	bbb21	657	komen	998
18	annalisa	1,577	agora	638	vandaag	970
19	mengonimarco	1,480	normani	631	anders	961
20	quando	1,416	arthurpicoli	617	waarom	946

Figure 28: The 20 most used words in the active networks.

BIBLIOGRAPHY

Allaway, Emily and Kathleen McKeown (2020). "Zero-shot stance detection: A dataset and model using generalized topic representations". In: *arXiv preprint arXiv:2010.03640*.

Alturayeif, Nora, Hamzah Luqman, and Moataz Ahmed (2023). "A systematic review of machine learning techniques for stance detection and its applications". In: *Neural Computing and Applications* 35.7, pp. 5113–5144.

Arnaboldi, Valerio, Marco Conti, Andrea Passarella, and Robin IM Dunbar (2017). "Online social networks and information diffusion: The role of ego networks". In: *Online Soc. Netw. Media* 1, pp. 44–55.

Arnaboldi, Valerio, Marco Conti, Andrea Passarella, and Fabio Pezzoni (2012). "Analysis of ego network structure in online social networks". In: *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*. IEEE, pp. 31–40.

— (2013). "Ego networks in twitter: an experimental analysis". In: *Proceedings IEEE INFOCOM*. IEEE, pp. 3459–3464.

Arnaboldi, Valerio, Andrea Passarella, Marco Conti, and Robin IM Dunbar (2015). *Online social networks: human cognitive constraints in Facebook and Twitter personal graphs*. Elsevier.

Augenstein, Isabelle, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva (2016). "Stance detection with bidirectional conditional encoding". In: *arXiv preprint arXiv:1606.05464*.

Barbieri, Francesco, Luis Espinosa Anke, and Jose Camacho-Collados (2021). "XLM-T: A multilingual language model toolkit for twitter". In: *arXiv preprint arXiv:2104.12250*.

Baumeister, Roy F, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D Vohs (2001). "Bad is stronger than good". In: *Review of general psychology* 5.4, pp. 323–370.

Biber, Douglas and Edward Finegan (1988). "Adverbial stance types in English". In: *Discourse processes* 11.1, pp. 1–34.

Boldrini, Chiara, Mustafa Toprak, Marco Conti, and Andrea Passarella (2018). "Twitter and the press: an ego-centred analysis". In: *Companion Proceedings of the The Web Conference 2018*, pp. 1471–1478.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini (2021). "The echo chamber effect on social media". In: *Proceedings of the National Academy of Sciences* 118.9, e2023301118.

Coe, Kevin, Kate Kenski, and Stephen A Rains (2014). "Online and uncivil? Patterns and determinants of incivility in newspaper website comments". In: *Journal of communication* 64.4, pp. 658–679.

Coleman, James S (1988). "Social capital in the creation of human capital". In: *American journal of sociology* 94, S95–S120.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov (2019). "Unsupervised cross-lingual representation learning at scale". In: *arXiv preprint arXiv:1911.02116*.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20, pp. 273–297.

Davis, James A (1967). "Clustering and structural balance in graphs". In: *Human relations* 20.2, pp. 181–187.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.

Dieng, Adji B, Francisco JR Ruiz, and David M Blei (2020). "Topic modeling in embedding spaces". In: *Transactions of the Association for Computational Linguistics* 8, pp. 439–453.

Dunbar, Robin I M (1998). "The social brain hypothesis". In: *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews* 6.5, pp. 178–190.

Dunbar, Robin IM (1992). "Neocortex size as a constraint on group size in primates". In: *Journal of human evolution* 22.6, pp. 469–493.

— (1993). "Coevolution of neocortical size, group size and language in humans". In: *Behavioral and brain sciences* 16.4, pp. 681–694.

Dunbar, Robin IM, Valerio Arnaboldi, Marco Conti, and Andrea Passarella (2015). "The structure of online social networks mirrors those in the offline world". In: *Social networks* 43, pp. 39–47.

Dunbar, Robin IM and Matt Spoors (1995). "Social networks, support cliques, and kinship". In: *Human nature* 6.3, pp. 273–290.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *kdd*. Vol. 96, pp. 226–231.

Ferrara, Emilio and Zeyao Yang (2015). "Quantifying the effect of sentiment on information diffusion in social media". In: *PeerJ Computer Science* 1, e26.

Fukunaga, Keinosuke and Larry Hostetler (1975). "The estimation of the gradient of a density function, with applications in pattern recognition". In: *IEEE Trans. on Inf. Theory* 21.1, pp. 32–40.

Gilbert, Eric and Karrie Karahalios (2009). "Predicting tie strength with social media". In: *Proceedings of the CHI*, pp. 211–220.

Goody, Esther (1972). "'Greeting','begging'and the presentation of respect". In: *The interpretation of ritual: Essays in honour of AI Richards*, pp. 39–72.

Gottman, John (1994). *Why marriages succeed or fail: And how you can make yours last*. Simon and Schuster.

Gottman, John, James Coan, Sybil Carrere, and Catherine Swanson (1998). "Predicting marital happiness and stability from newlywed interactions". In: *Journal of Marriage and the Family*, pp. 5–22.

Gottman, John, Howard Markman, and Cliff Notarius (1977). "The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior". In: *Journal of Marriage and the Family*, pp. 461–477.

Granovetter, Mark S (1973). "The strength of weak ties". In: *American journal of sociology* 78.6, pp. 1360–1380.

Grootendorst, Maarten (2022). "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". In: *arXiv preprint arXiv:2203.05794*.

Grover, Aditya and Jure Leskovec (2016). "node2vec: Scalable feature learning for networks". In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864.

Hart, Betty and Todd R Risley (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing.

Hassan, Ahmed, Amjad Abu-Jbara, and Dragomir Radev (2012). "Extracting signed social networks from text". In: *Workshop Proceedings of TextGraphs-7*, pp. 6–14.

Heatherly, Kyle A, Yanqin Lu, and Jae Kook Lee (2017). "Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites". In: *New Media & Society* 19.8, pp. 1271–1289.

Heider, Fritz (1946). "Attitudes and cognitive organization". In: *The Journal of psychology* 21.1, pp. 107–112.

Hill, Russell A and Robin IM Dunbar (2003). "Social network size in humans". In: *Human nature* 14.1, pp. 53–72.

HuggingFace (2022). *Sentiment Analysis Model*. URL: `https://huggingface.co/sbcBI/sentiment_analysis_model` (visited on 03/03/2023).

Hutchens, Myiah J, Vincent J Cicchirillo, and Jay D Hmielowski (2015). "How could you think that?!?!: Understanding intentions to engage in political flaming". In: *New media & society* 17.8, pp. 1201–1219.

Hutto, Clayton and Eric Gilbert (2014). "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In: *Proceedings of ICWSM*. Vol. 8, pp. 216–225.

Javari, Amin and Mahdi Jalili (2014). "Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links". In: *ACM TIST* 5.2, pp. 1–19.

Jenks, George F (1967). "The data model concept in statistical mapping". In: *International yearbook of cartography* 7, pp. 186–190.

Kadivar, Jamileh (2017). "Online radicalization and social media: A case study of Daesh". In: *International Journal of Digital Television* 8.3, pp. 403–422.

Khiabani, Parisa Jamadi and Arkaitz Zubiaga (2023). "Few-shot learning for cross-target stance detection by aggregating multimodal embeddings". In: *IEEE Transactions on Computational Social Systems*.

Lau, Jey Han, David Newman, and Timothy Baldwin (2014). "Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality". In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 530–539.

Leskovec, Jure, Daniel Huttenlocher, and Jon Kleinberg (2010a). "Predicting positive and negative links in online social networks". In: *Proceedings of WWW*, pp. 641–650.

— (2010b). "Signed networks in social media". In: *Proceedings of the CHI*, pp. 1361–1370.

Li, Yingjie, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea (2021). "P-stance: A large dataset for stance detection in political domain". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 2355–2365.

Liu, Bing (2012). "Sentiment analysis and opinion mining". In: *Synthesis lectures on human language technologies* 5.1, pp. 1–167.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019). "Roberta: A robustly optimized bert pretraining approach". In: *arXiv preprint arXiv:1907.11692*.

MacQueen, James (1967). "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 14. Oakland, CA, USA, pp. 281–297.

Maniu, Silviu, Talel Abdessalem, and Bogdan Cautis (2011). "Casting a web of trust over wikipedia: an interaction-based approach". In: *Comp. proceedings of WWW*, pp. 87–88.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Miritello, Giovanna, Esteban Moro, Rubén Lara, Rocío Martínez-López, John Belchamber, Sam GB Roberts, and Robin IM Dunbar (2013). "Time as a limited resource: Communication strategy in mobile phone networks". In: *Social networks* 35.1, pp. 89–95.

Mohammad, Saif, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry (2016). "Semeval-2016 task 6: Detecting stance in tweets". In: *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pp. 31–41.

Mowen, John C and Stephen W Brown (1981). "On explaining and predicting the effectiveness of celebrity endorsers." In: *Advances in consumer research* 8.1.

Munger, Kevin, Patrick J Egan, Jonathan Nagler, Jonathan Ronen, and Joshua Tucker (2022). "Political knowledge and misinformation in

the era of social media: Evidence from the 2015 UK election". In: *British Journal of Political Science* 52.1, pp. 107–127.

Nguyen, C Thi (2020). "Echo chambers and epistemic bubbles". In: *Episteme* 17.2, pp. 141–161.

Nguyen, Dat Quoc, Thanh Vu, and Anh Tuan Nguyen (2020). "BERTweet: A pre-trained language model for English Tweets". In: *arXiv preprint arXiv:2005.10200*.

Oates, Joan (1977). "Mesopotamian social organisation: Archaeological and philological evidence". In: *The evolution of social systems*, pp. 457–485.

Ollivier, Kilian, Chiara Boldrini, Andrea Passarella, and Marco Conti (2022). "Structural invariants and semantic fingerprints in the" ego network" of words". In: *arXiv:2203.00588*.

Ostrom, Elinor (2003). "Toward a behavioral theory linking trust, reciprocity, and reputation." In: *Trust and reciprocity: Interdisciplinary lessons from experimental research*, pp. 19–79.

Paraskevopoulos, Pavlos, Chiara Boldrini, Andrea Passarella, and Marco Conti (2021). "The academic wanderer: Structure of collaboration network and relation with research performance". In: *Applied Network Science* 6, pp. 1–35.

Rababah, Mahmoud Ali and Nibal Abd Alkareem Malkawi (2012). "The linguistic etiquette of greeting and leave-taking in Jordanian Arabic". In: *European Scientific Journal* 8.18.

Reimers, Nils and Iryna Gurevych (2019). "Sentence-bert: Sentence embeddings using siamese bert-networks". In: *arXiv preprint arXiv:1908.10084*.

— (2020). "Making monolingual sentence embeddings multilingual using knowledge distillation". In: *arXiv preprint arXiv:2004.09813*.

Rosenthal, Sara, Noura Farra, and Preslav Nakov (2019). "SemEval-2017 task 4: Sentiment analysis in Twitter". In: *arXiv preprint arXiv:1912.00741*.

Rozin, Paul and Edward B Royzman (2001). "Negativity bias, negativity dominance, and contagion". In: *Personality and social psychology review* 5.4, pp. 296–320.

Schöne, Jonas Paul, Brian Parkinson, and Amit Goldenberg (2021). "Negativity spreads more than positivity on Twitter after both positive and negative political situations". In: *Affective Science* 2.4, pp. 379–390.

Shi, Guodong, Alexandre Proutiere, Mikael Johansson, John S Baras, and Karl H Johansson (2016). "The evolution of beliefs over signed social networks". In: *Operations Research* 64.3, pp. 585–604.

Statistica (2022). *Number of monetizable daily active X (formerly Twitter) users (mDAU) worldwide from 1st quarter 2017 to 2nd quarter 2022*. URL: https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide (visited on 05/15/2024).

Statistica (2024). *Leading countries based on number of X (formerly Twitter) users as of April 2024*. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries (visited on 05/17/2024).

Sun, Renjie, Qiuyu Zhu, Chen Chen, Xiaoyang Wang, Ying Zhang, and Xun Wang (2020). "Discovering cliques in signed networks based on balance theory". In: *Database Systems for Advanced Applications: 25th International Conference, DASFAA 2020, Jeju, South Korea, September 24–27, 2020, Proceedings, Part II 25*. Springer, pp. 666–674.

Sun, Shiliang, Chen Luo, and Junyu Chen (2017). "A review of natural language processing techniques for opinion mining systems". In: *Information fusion* 36, pp. 10–25.

Sutcliffe, Alistair, Robin Dunbar, Jens Binder, and Holly Arrow (2012). "Relationships and the social brain: integrating psychological and evolutionary perspectives". In: *British journal of psychology* 103.2, pp. 149–168.

Swire, Briony, Adam J Berinsky, Stephan Lewandowsky, and Ullrich KH Ecker (2017). "Processing political misinformation: Comprehending the Trump phenomenon". In: *Royal Society open science* 4.3, p. 160802.

Tacchi, Jack, Chiara Boldrini, Andrea Passarella, and Marco Conti (2022). "Signed ego network model and its application to Twitter". In: *IEEE BigData 2022*.

— (2023). "Cultural Differences in Signed Ego Networks on Twitter: An Investigatory Analysis". In: *Companion Proceedings of the ACM Web Conference 2023*, pp. 1039–1049.

— (2024). *Keep Your Friends Close, and Your Enemies Closer: Structural Properties of Negative Relationships on Twitter*. arXiv: 2401.16562 [cs.SI].

Tang, Jiliang, Yi Chang, Charu Aggarwal, and Huan Liu (2016). "A survey of signed network mining in social media". In: *ACM Computing Surveys (CSUR)* 49.3, pp. 1–37.

Toprak, Mustafa, Chiara Boldrini, Andrea Passarella, and Marco Conti (2021). "Structural Models of Human Social Interactions in Online Smart Communities: the Case of Region-based Journalists on Twitter". In: *arXiv preprint arXiv:2110.01925*.

— (2022a). "Harnessing the Power of Ego Network Layers for Link Prediction in Online Social Networks". In: *IEEE Trans. Comput. Soc. Sys.*

— (2022b). "Journalists' ego networks in Twitter: Invariant and distinctive structural features". In: *Online Social Networks and Media* 30, p. 100207.

— (2023). "Harnessing the Power of Ego Network Layers for Link Prediction in Online Social Networks". In: *IEEE Trans. Comput. Soc. Syst.* 10.1, pp. 48–60.

Traag, V. A. and Jeroen Bruggeman (2009). "Community detection in networks with positive and negative links". In: *Physical Review E* 80.3. ISSN: 15393755.

UKinbound (2020). *List of MP Twitter Accounts*. URL: https://www.ukinbound.org/resources/list-of-mp-twitter-accounts (visited on 03/03/2022).

Wei, Penghui and Wenji Mao (2019). "Modeling transferable topics for cross-target stance detection". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1173–1176.

Wenzek, Guillaume, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave (2019). "CCNet: Extracting high quality monolingual datasets from web crawl data". In: *arXiv preprint arXiv:1911.00359*.

Xu, Chang, Cécile Paris, Surya Nepal, and Ross Sparks (2018). "Cross-target stance classification with self-attention networks". In: *arXiv preprint arXiv:1805.06593*.

Yuan, Weiwei, Jiali Pang, Donghai Guan, Yuan Tian, Abdullah Al-Dhelaan, and Mohammed Al-Dhelaan (2019). "Sign Prediction on Unlabeled Social Networks Using Branch and Bound Optimized Transfer Learning". In: *Complexity* 2019. ISSN: 10990526.