Friend or Foe: Linguistic Signals of Community and Impoliteness in Online Anonymous Interactions

Teona Banu
Trinity College Dublin
banut@tcd.ie

Vivek Bhadula
Trinity College Dublin
bhadulav@tcd.ie

Shuncheng Cai
Trinity College Dublin
cais@tcd.ie

Xiaohan Shi
Trinity College Dublin
shixi@tcd.ie

Mengqi Wang
Trinity College Dublin
wangm8@tcd.ie

Abstract

We investigate the effects of user anonymity, in the form of user identity persistence and message persistence, on linguistic signals of community and politeness. We assemble a new corpus and make it available for use in future research. We find that user anonymity is negatively correlated with both politeness and community, with identity persistence having a larger effect on politeness, and message persistence having a larger effect on community.

Keywords: anonymity, politeness, toxicity, signals of community, online platforms, text analysis, computational linguistics

1 Introduction

1.1 Background

In the age of Internet communication, the pros and cons of anonymity have become increasingly controversial. On the one hand, some believe that anonymous communication can promote constructive discussion by removing personality and personalization from interactions. On the other hand, some believe that anonymous communication can easily lead to abhorrent behaviour. Harmful practices have been described on some online platforms that support anonymous communication (Papasavva et al., 2020). At the same time, certain behaviours seen as unquestionably harmful in some contexts may be signals of community favouring in others, depending on the community used (Locher and Watts, 2008). Online platforms that offer to build a lasting personal image have better mechanisms for building community than those where anonymity is the norm. Therefore, it is necessary to investigate whether there are differences in language signals, including community and polite or impolite language signals, between platforms that support anonymous communication and platforms that support building durable identities. Such research can contribute to a better understanding of the role of anonymity and enduring identity in shaping online discourse and guide the development of more effective online communication policies and practices.

1.2 Research Topic

Our research topic investigates language signals in anonymous online interactions, particularly community and impolite language signals. Our research aims to explore whether there are differences in community and polite or impolite language signals between online platforms that support anonymous communication and those that support the construction of lasting identities. By analyzing the language use of these two types of platforms, we hope to gain a deeper understanding of the impact of anonymity and enduring identity on online discourse. Through this study, we aim to contribute to the ongoing discussion about the role of anonymity in shaping online interactions and facilitating constructive discussion while minimizing harmful behaviours.

1.3 Research Questions

There is currently no consensus on the effects of anonymity in communication. Toxic behaviour has been observed on some platforms that support user anonymity (Papasavva et al., 2020), but other authors have found that anonymity or unidentifiability may not be the main factor at play and that other features of online communication, such as lack of eye contact, may play a bigger role in online toxicity (Lapidot-Lefler and Barak, 2012). For this reason, we believe it would be worthwhile contrasting different online communities in terms of the linguistic signals of politeness and toxicity while paying special attention to the extent to which user accounts can be traced back to the person and the stability of the conversations. Thus, our first research question will be

Do linguistic signals of politeness differ across online platforms employing different degrees of anonymity?

Next, we want to look at the effects of anonymity on establishing a community. Evidence suggests that anonymity might be compatible with (Postmes et al., 2001), if not nurturing (Smith et al., 2007), adherence to group norms. It has been observed that anonymity leads to an increased propensity to share information and ideas. However, this effect differs across demographics Durant et al. (2002) and does not inherently lead to better quality results in goal-oriented interactions Connolly et al. (1990). Nevertheless, the potential positive effects of anonymity on interaction lead to our second research question:

Do linguistic signals of community differ across online platforms employing different degrees of anonymity?

We do not expect platforms where anonymity leads to toxicity automatically to have a low sense of community. On the contrary, it might be the case that what from the outside is perceived as anti-social behaviour is, in fact acceptable or even encouraged for group members. This is based on the idea of Locher and Watts (2008) that when considering the intention and the effect of communications, impoliteness is not an objective measure and instead is determined by the group as part of the "frame". The results gathered from looking to answer the first two research questions will give rise to a discussion which seeks to answer our third and final research question:

Is there a link between politeness and community on online platforms with respect to user anonymity?

1.4 Hypotheses

We are looking at four social media platforms - Tumblr and Reddit, which both have persistent user identities, Twitch, which has persistent user identities but a strong collective identity and impermanent messages, and 4chan, which has impermanent user identities but somewhat persistent messages. We hypothesise that Tumblr and Reddit will not differ significantly, as the topic of discussion is narrowed to a single topic for all four platforms - video games - and the two platforms have the same type of user identity. We also hypothesise that higher levels of anonymity will lead to lower levels of politeness and higher levels of toxicity. We especially expect to find high levels of toxicity on 4chan but also high levels of community on Twitch. We expect to find at least some cases where community and politeness are negatively correlated, but it will be interesting to see the effect of anonymity on this.

1.5 Contributions

Our main findings help decipher the relationship between anonymity, community, and politeness. We find that both politeness and community are higher when persistent user identities and messages are present and that the lack of persistent messages most strongly impacts community, while the lack of persistent user identities most strongly impacts politeness. We confirm that the phenomenon of a stronger community leading to less politeness does occur. We also assemble a brand new data set which can be accessed and used for future research into text analysis of online content. [1]

The rest of this paper is structured as follows: section 2 offers an overview of previous literature; Section 3 introduces the corpus that has been constructed for this study; Section 4 details the methodology and tools that were used, while Section 5 presents the results obtained; section 6 offers a discussion of these results and their implications, as well as concluding remarks and suggestions for future research.

[1] The corpus as well as the code we wrote, is available at: https://github.com/angryathena/Friend-or-Foe-Linguistic-Signals-of-Community-and-Impoliteness-in-Online-Anonymous-Interactions.git

2 Literature Review

The literature on the topic of anonymous communication on online platforms reveals a range of perspectives and findings. We have examined the ways in which anonymity affects communication, including how it impacts identity expression, social interaction, and the formation of online communities. The literature suggests that anonymity can have both positive and negative effects on online communication, depending on a variety of factors such as the nature of the platform, the goals of the users, and the norms and regulations that govern online behaviour.

2.1 Anonymity

The article from Vogel (2014) explores the possible advantages and disadvantages of anonymity in online communication. According to some experts, anonymity can promote creativity and encourage collaborative thinking, but it can also create racial and gender tensions, violence, and selfishness. The author proposes FreeSpeech, a brand-new platform that permits anonymous speech while simultaneously removing offensive and unlawful content using content analysis and machine learning. The author does agree that it can be challenging to identify all the bad material and that users may figure out a way to go around the system. Despite these difficulties, the author argues that anonymity is very important for free speech since it shields people from societal, political, and financial pressures while enabling wider idea expression. The author contends that while steps can be taken to lessen the drawbacks of anonymity, it should not be completely abandoned.

Zhang and Kizilcec (2014) investigates how sharing behaviour on social media, particularly when deciding between public and anonymous sharing, is influenced by content that is contentious and socially accepted. Social media platforms provide connectedness and information exchange, yet there are worries about privacy and the release of sensitive data. With interface designs, researchers try to balance users' social behaviour and privacy concerns. While anonymity can make emotions like anger and encourage inappropriate sharing of explicit content, public sharing provides a psychologically fulfilling experience. According to the study, social media content familiarity had a substantial impact on sharing behaviour, with fresh content having a higher likelihood of being shared anonymously. Social media interface designs should provide anonymity choices for users who seek to conceal their identity because users favour anonymous sharing over public sharing. Being surrounded by people spreads blame and promotes anonymity, which encourages more inappropriate behaviour. The study's overall conclusion is that privacy issues should be considered when developing social media user interfaces so that users can contribute content anonymously if they so choose.

2.2 Politeness and Toxicity

The politically incorrect sector of the anonymous 4chan website, which is well-known for its politically wrong, racist, sexist, and hateful comments, is the subject of the paper by Papasavva et al. (2020). The authors present a dataset of more than 3.3 million posts and 134.5 million comments for academics interested in investigating 4chan or other online identities after using natural language processing techniques to examine over 12 million posts and 6 million comments. The crawling data, which covers trending subjects like immigration, Islam, black people, women, nationalism, and the rising use of acronyms and code words, is described in depth by the authors, along with its origin and organizational structure. Additionally, the dataset's statistical and content analysis is included in the report, revealing the most prevalent patterns and toxicity levels. The authors make use of a variety of technologies, including Latent Dirichlet Allocation for modelling, TF-IDF for natural language processing, API for toxicity assessment, and spaCy for entity assessment. The authors want to support and promote studies in several fields, such as the effects of forum anonymity, natural language processing, and human behaviour analysis. Managing communities on 4chan is challenging due to the site's anonymity, and as the site's post count grows over time, so does its influence.

The analysis by Young et al. (2018) in this paper looks at 993 question and response binary groups in order to analyze the discourse tactics utilized in offensive postings, answers, and bystander comments on social media sites. The attackers in this study are essentially fully anonymous. Only 2.8% of the aggressors included their profile name in the initial aggressive posts, which were almost all made anonymously. Posters with inflammatory content make fun of profile owners based on their age, perceived physical attractiveness, social position, and mental instability.

Dillon et al. (2015) examined the tone, civility, and politeness of the online comments on four Ohio newspapers. They discovered that anonymity significantly influenced the tone of the remarks, with anonymous individuals more inclined to behave impolitely and uncivilly. The findings also revealed that people who posted comments through anonymous Facebook users were less courteous and civil than those who posted comments using their own Facebook identities. Compared to non-political news, political news stories garnered less respectful and civil comments.

The relationship between relationship work and rudeness in interpersonal communication is examined by Locher and Watts (2008). They define impoliteness as speech that violates social norms, and related work as the verbal and nonverbal behaviours employed to sustain positive connections. The authors stress that maintaining healthy relationships can be facilitated by adhering to reasonable expectations and that these expectations are continually subject to cultural change. The writers also cover the topic of applying behavioural standards to the internet, where users are more likely to transgress norms because of anonymity, disparities in social norms, and individual preferences. Negotiating and settling disagreements in these circumstances requires relationship work; else, people may become offended. To show instances of norm violations, the writers examine a political interview video. The host continually interrupts the guest during the interview and uses derogatory language to challenge the visitor's perspective, which is disrespectful to the guest and goes against conversational etiquette. This behaviour not only disrupts the conversation's environment but also has an impact on the participants' interactions with one another. According to the authors, maintaining excellent relationships in interpersonal communication requires relational work, and being disrespectful can damage relationships. They contend that more study is required to fully comprehend the function of connection labour in controlling impoliteness in various communication circumstances.

2.3 Community

Online communities with a lot of people and messages tend to have high cohesion, as indicated by the use of "we" phrases, and fairly high sociability, as indicated by the number of emotions communicated by users, according to McEwan (2016), who looked at the communication patterns of these groups. High cohesion, high emotionality, and high social word usage are associated with high interactivity, although these characteristics do not rise linearly with interaction. Instead, they rise as interaction levels reach a critical level.

In their essay, Code and Zap (2009) investigate the role that social identities play in the emergence and expansion of online communities. They go on to say that people are encouraged to express themselves freely and form connections with other people who share their beliefs because of the anonymity of internet communication. Participation in social groups is the basis of individuals' social identities, and because of the prevalence of online social media, it is now simpler than ever to locate and get in touch with people who have similar interests. People categorize themselves and other people to form social identities, and the degree to which a person may strengthen their social identity depends on how well they can identify with that group. The study shows that social networks develop constantly as a result of visualization tools and advertising, and as users become accustomed to the platform, it becomes an essential part of their lives. In general, social media platforms give users a platform to explore and express their social identities, and understanding intergroup linkages is crucial for linking users to their social groups.

3 Data

We want to look at multiple online platforms with different implementations of user anonymity. Facebook has a very low degree of anonymity as users conventionally use their own names and pictures. Users often use Facebook to connect with people they know in real life such as friends, family, and coworkers.

We took note of the work of Khalid and Srinivasan (2020) which showed that linguistic markers differ not only across platforms but also across boards and topics within platforms. Because of this, we are using, for each platform studied, datasets comprised of users discussing the same topic. We chose the topic of "video games" for a list of reasons. First, it is a neutral subject about which people tend to feel passionate enough to share their opinions freely, but not enough to forgo their inhibitions and perform out-of-character verbal assaults on those who disagree with them. This is in direct contrast to politically-charged topics explored by Papasavva et al. (2020). Furthermore, video games have become a popular hobby for people of all ages, especially in recent times after people were forced to isolate and turned to long-distance options of getting in touch with others and performing activities together. Thus, we are likely to see a not-unrepresentative sample of what the general online population might act like. Finally, we decided early on that we wanted to include Twitch due to its chat message ephemerality, which resembled one of the important qualities listed by Vogel (2014) as contributing to anonymity in online discussions. Since Twitch is most commonly used for video game streams, we decided to turn our biggest constraint into the deciding factor for the user discussion topic.

```
[
  {
    "thread": "<p>\u201cI wish I could make a game but get other people to do all the art and writi
    "comments": [
      "How do I find that, especially as a minor",
      "People don't quit bad jobs, they quit bad managers. I have quit at least one PM.",
      "This is true in any field and as a people manager, I can say that you must also be comfort
      "I'm also a little freak who enjoys organizing work stuff! It's not boring to me, it's fun
      "Oh, i wish i could do that."
    ]
  },
  {
    "thread": "<p>I love you Stardew Valley I love you Stray I love you Untitled Goose Game I love
    "comments": [
      "Check out A Short Hike, it's a small and lovely game. It helped me get out of a really bad
      "I need to finish Lost Ember before I get Stray, but yeah.",
      "BATIM & BATDR <]",
      "I love you Project Wingman",
      "Lil Gator Game"
    ]
  },
  ...
]
```

Figure 1: Dataset format exemplified by two threads from the Tumblr dataset

All of our datasets have been formatted in the same way. Each platform we are investigating corresponds to a JSON file with the format shown in figure 2. The ”thread” key points to the initial post, or thread topic. The ”comments” key points to the list of sub-comments comprising the discussion under the thread.

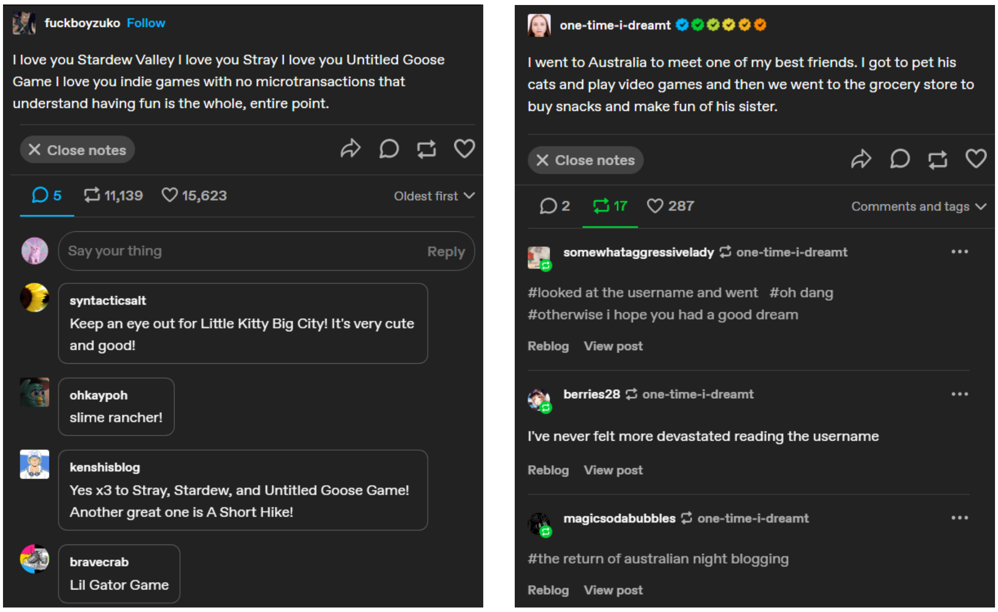
3.1 4chan

Figure 2: 4chan post with some replies

We scraped 100 recent posts from the /v/ - Video Games board. The code was based on the work of boilingpenguin (2018) available on their GitHub. The dataset we created includes, for each post, the original post, or the thread topic, and a list of replies. Figure ?? shows what a 4chan post looks like. The main comment and often a figure are shown at the top and are followed by a list of replies. Under a thread, users can choose to reply to the original poster or to another reply post. 4chan does not support persistent user identities or accounts. Instead, most users make posts and reply under the name

"Anonymous". The platform allows users to input a name when posting, but this is very rarely used, with the "Anonymous" identity being so entrenched in the 4chan culture that it is common for users to affectionately call each-other "anon".

3.2 Tumblr



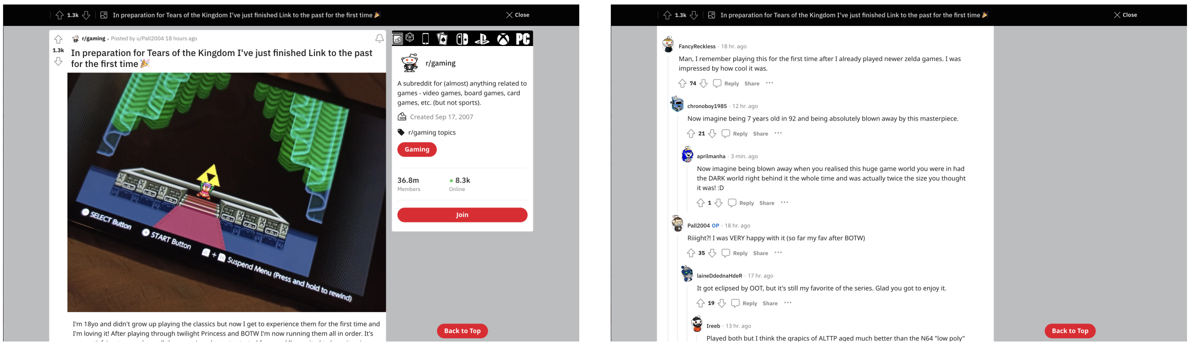
(a) Replies (b) Reblogs

Figure 3: Two Tumblr posts with some reblogs and replies respectively

The data was gathered by us using the Tumblr API. The dataset includes, for each post, the original post, or thread topic, and a list of comments. We consider "comments" to mean both replies and reblogs. Replying to a Tumblr post is a feature that most closely resembles a typical comment section, with users contributing to the post's reply thread. However, since Tumblr originated as a microblogging platform, its oldest and best-known feature is the "Reblog" option, which adds the post to the user's blog, allowing the user to make an addition via text or tags. The difference between text and tags is fundamental for the culture of Tumblr, with text reblogs becoming their own post which can later be reblogged by others, and tags only showing up on the blog of the reblogging user and on the reblog thread. Because of this, when scraping the "comments" under a post, we are scraping replies, text additions from reblogs and reblog tags. While the purposes of each type of "comment" are considered distinct on the platform, we do not need to make any distinction for the purposes of our research. Figure 3 shows two Tumblr posts with visible (a) replies and (b) reblogs including both text and tags. Tumblr has persistent user identities but they are most often not easily traceable back to the person as it is unusual and discouraged to share real information, especially pictures and names. Users often use this platform to maintain an online identity that is closer to their "true" personality and interests. This can be seen in long-standing inside jokes about how users are forbidden from explicitly bringing up Tumblr in face-to-face conversations and must instead rely on coded language. This idea gained enough traction among the users to birth official Tumblr merchandise which gives a nod to this coded language.

3.3 Reddit

The data regarding Reddit was obtained by scraping the subreddit section of r/gaming. The code used for scraping was based on the work pistocop (2022) published on GitHub. We collected the top 100 posts from the "hot" category of the r/gaming subreddit, which included the original content or topic of each post, as well as the list of replies. As illustrated in the figure4, a typical post in the r/gaming subreddit includes a topic and a comment section where users can respond to the topic or any comment. All these conversations were collected in our data collection process. The general culture of Reddit is anonymous, which allows users to post and comment without revealing their real identities. Due to the many challenges surrounding morality and politeness, moderators will establish rules and promptly remove toxic comments in order to strive for a balance between freedom of speech and responsible behaviour.



(a) Topic (b) Replies

Figure 4: A Reddit posts with some replies

3.4 Twitch

We are using a dataset created by Kim (2019). It is a collection of chat logs of 2,162 Twitch streaming videos corresponding to 52 streamers. The time period of the target streaming videos is from 2018-04-24 to 2018-06-24. We are using the first two streams for each of the streamers in order to get a dataset of a similar size to the other platforms. In contrast to the other platforms studied, Twitch does not have an initial post. Instead, the list of comments corresponds to live video stream chat messages. The "thread" key points to the video id, and the "comments" key points to the list of messages in the chat. Twitch chats are characterised by the rapid sharing of text messages and emojis/stickers. Messages in the chat are closer to texting or real-time conversations than to posts that are meant to be recorded for posterity. Figure 5 shows a snapshot of a Twitch chat. This particular example shows a practice known as "chat spamming" where viewers post multiple consecutive messages comprised of a single word or sticker repeated many times. This common practice is a testament to the ephemerality of chat messages - not only do chats disappear after a stream ends, but the debit of incoming messages can be so high as to cause the chat screen to change completely multiple times per second. Twitch requires users to sign up and create a persistent identity. That being said, due to the rapid sharing of chat messages, it is rare for users to stand out or even be recognised as individuals. This is seen in the way streamers rarely refer to individual users, but very often refer to all their viewers collectively by affectionately calling them "Chat".

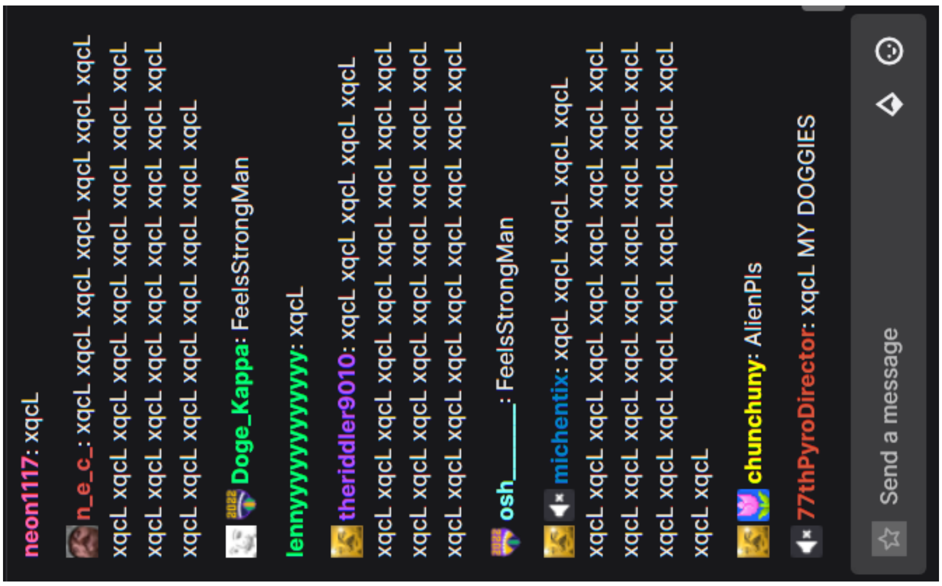


Figure 5: Snapshot of a Twitch chat

	Platform	Min	Median	Mean	Max	StDev
Comments per Thread	Tumblr	1	4	6	46	7
	Reddit	1	8	80	1644	276
	Twitch	1011	20985	29241	145689	29175
	4chan	1	35	100	484	126
Words per Comment	Tumblr	1	12	20	99	21
	Reddit	1	11	19	557	29
	4chan	1	13	21	347	27
	Twitch	1	2	5	156	7

Table 1. *Corpus descriptive statistics*

Table 1 shows some descriptive statistics for the corpus. Reddit and 4chan are similar in terms of comments per thread, with Tumblr having much fewer. Tumblr, Reddit, and 4chan are very similar in terms of words per comment, with almost exactly the same median and mean. Twitch is an outlier in that it has a lot more comments per thread but they tend to be a lot shorter. However, only 100 comments are sampled for each thread, which somewhat evens out the sizes of the threads.

4 Methodology

4.1 Preprocessing

After getting the file in JSON format, we performed data pre-processing. Specifically, we performed data cleaning to remove invalid information and noise from it in order to obtain cleaner and more reliable data. The specific steps include:

- Reads the data from the original file, this includes the "thread" and the corresponding "comments"
- Define a cleaning function that uses regular expressions to remove HTML tags, URLs, invalid information, invalid symbols, labels, and numbers with more than five consecutive digits, etc.
- Iterate through each entry, applying the purge function to the text content in the "thread" and "comments" fields
- Store the processed data in a new file

We then created approximately 100 strings for each platform, each representing one "thread". For the platforms with fewer comments, the threads were generated by concatenating all comments, including the original post. The chat transcripts of the Twitch dataset contained too many comments for integral threads to be feasible. We decided to generate the strings using a sample of chat messages. We believe different points of a stream are likely to generate different types of messages - for example, messages toward the end of a stream are likely to have an inflated politeness value with viewers thanking the streamer. For this reason, we wanted to avoid sampling a specific section of the chat transcript and instead took random samples from the entire set of messages.

4.2 Feature extraction

We are extracting 13 features from the data and then computing 3 aggregate features from the initial 13, to a total of 16 features. Thus, each platform will have a list of approximately 100 observations with 16 features each.

4.2.1 Community

We base some of our measures of community on the work of McEwan (2016). We measure 'we-ness' to gauge cohesiveness, past and future verb tenses for group stability, and emotions for sociability and trust. We also use the F score defined by **??** to measure contextuality. We are using TreeTagger to extract morphological features from each thread. We are using the BNC tagset, which is intended for use on corpora of both written and spoken English. Thus, it can recognise informal and colloquial use, which our dataset is likely to be filled with.

To measure we-ness, we define a we-count method which returns, for each thread, the occurrences of the words "we, us, ours, ourselves, ourself" as a fraction of the total number of words in the thread.

We use a similar procedure for verb tenses - returning the fraction of verbs that are at past and future tense out of all the verbs in the thread. To be more specific, we are looking at each individual predicate, by only counting one of the verbs making up a continuous or perfect verb tense. We define our own list of parts of speech which we search for in our data. Past tense indicators are taken to be the past tense forms of lexical verbs and of the verbs "to be", "to do", and "to have" (danced, was, did, had etc). To count present tense references, we count the number of present tense forms of the verb "to be" (am, are is etc), the finite base form of "to do", "to have", and lexical verbs (do, have, dance), and -s form of "to do", "to have", and lexical verbs (does, has, dances). Finally, for future references, we count the occurrences of the modal auxiliary verbs "will" and "shall". All other verb forms are considered to go alongside one of these verbs to form a singular temporal reference and thus are not counted. We acknowledge two main issues with our approach. First, we do not count the number of past participle verb forms (been, done, forgotten) which appear in continuous verb forms. Thus, verbs at the present continuous and present perfect continuous tense are counted as part of the present tense group. It is debatable whether this should be the case for our purposes. Even though "I have danced" is grammatically a present tense structure, it contains a logical reference to the past. However, we decided against adjusting for this as it would have caused issues with the way other counters behave, such as causing past (perfect) continuous verbs to be counted as two references to the past and the structure "I will have danced" to be counted as one future and one past reference. The second issue is that the only future tense we can count are constructions using "will", "'ll", or "shall". Structures such as "I am going to dance" or "I am dancing tomorrow" are counted as present references even though logically a human would interpret them as future references.

The F score is computed by counting the occurrence of all parts of speech in the threads, summing the non-deictic words and subtracting deictic words. We did not stray from the method introduced by **?**, but one possible point of improvement in our case could have been counting "Unclassifed" words as contributing to higher contextuality. Even though the BNC tagset contains many colloquial terms, new "words" pop up every day on social media. Especially on Twitch, figure 5 shows "chat" using a seemingly meaningless word that carries meaning only to the community.

Negative emotion and positive emotion are meaningful numerical features obtained by quantifying and normalizing the scores of negative emotion and positive emotion in the text. For example, counting the frequency with which negative emotion words are used can yield a quantitative "negative emotion usage" characteristic to assess the level of negative emotion in a topic; The analysis and evaluation of tone results in a quantitative "tone appropriateness" characteristic, which can be used to assess the politeness of tone in a topic. Similarly, the frequency of use of non-offensive language and the appropriateness of emotional expression can be quantified and analyzed to obtain the corresponding politeness characteristics, which can be used to evaluate the politeness of the topic. By acquiring these features, we can have a deeper understanding of the emotional tendencies in the text, and thus provide references for further analysis and decision-making.

We used Empath, an open-source sentiment analysis tool, to do sentiment analysis on these texts and extract negative and positive sentiment scores.

The aggregate feature "Emotion" is computed by summing the positive emotion and negative emotion scores. Normally, when performing sentiment analysis, positive and negative emotions would have different signs and a sentence where the user expresses one positive emotion and one negative emotion would be counted as overall neutral. However, because we are computing the trust and sociability of users by using their propensity to share emotions, the same sentence would be computed by us as two counts of emotion sharing. Another aggregate feature, "Community" sums each individual community-related feature (after normalisation) and subtracts the formality score.

4.2.2 Politeness

Politeness features are meaningful and numerical features obtained by quantifying and normalising certain polite behaviours in the process of quantifying and analysing polite behaviours. For example, statistics and calculations on the frequency of use of polite words can yield a quantitative 'politeness usage rate' feature, which can be used to assess the politeness of a topic; analysis and evaluation of the appropriateness of tone can yield a quantitative 'tone appropriateness "The analysis and evaluation of tone of voice give a quantitative 'tone of voice' feature, which is used to assess politeness in the subject matter. Similarly, the frequency of use of non-offensive language and the appropriateness of emotional expressions can be quantified and analysed to obtain a corresponding politeness characteristic for assessing politeness in the subject matter.

We are also computing occurrences of outright impoliteness by considering various scores of (seemingly) antisocial behaviour - (severe) toxicity, identity attack, insults, profanity, threats, and inflammatory comments. To do so, we are using Perspective API and output the results for each relevant feature from each thread.

The aggregate feature "Politeness" is computed by subtracting all the (normalised) toxicity features from the (normalised) politeness feature.

4.2.3 Normalisation and descriptive statistics

After extracting the individual features, we are applying min-max normalisation to ensure all of the values are between 0 and 1. This ensures that when computing the aggregate features, all of the individual features have equal weights. After computing the aggregate features, we compute z-scores for them as these values will be used in the regression analysis.

We then compute, for each feature, the minimum, maximum, mean and standard deviation. Following the min-max normalisation process, the minimum values will always be larger or equal to 0, with at least one platform having a minimum equal to 0, and the maximum values will be smaller or equal to 1, with at least one platform having a maximum value of 1. The z-scores will always

4.3 Analysis

4.3.1 Exploratory analysis

We first begin by performing a few tests on our data to see what assumptions we can proceed with. We compute the correlation between the features using Python and plot it using a heatmap.

We then use a Shapiro-Wilk test (Shapiro and Wilk, 1965) to see if our features are normally distributed. For this, we use the Scipy package in Python. The null hypothesis being tested is that the sample of observations from each feature follows a normal distribution. The test statistics being computed for each platform-feature combination are shown in formula 1

$$W_{pf} = \frac{(\sum_{i=1}^{n_p} a_{ipf} x_{(i)pf})^2}{\sum_{i=1}^{n_p} (x_{ipf} - \bar{x_{pf}})^2} \tag{1}$$

where n_p is the number of threads from platform p, x_{ipf} is the value of the feature f for thread i from platform p, $x_{(i)pf}$ is the i^{th} order statistic of feature f for platform p, $\bar{x_{pf}}$ is the average value of feature f for platform p. A small p-value will lead to the rejection of the null hypothesis and a conclusion that the data is not normally distributed and we will have to proceed with non-parametric tests which do not rely on the assumption that the data is normally distributed.

4.3.2 Tests

We are using the non-parametric Kurskal-Wallis (Kruskal and Wallis, 1952) test, available through the scipy.stats package, to determine whether the platforms differ from one another in terms of each feature. The validity of this test relies upon a series of assumptions about the dataset (Laerd Statistics, 2018). The test assumes that the dependent variable is a continuous variable. This assumption is fulfilled in our data by the features being tested which all consist of continuous measures between 0 and 1. The independent variable is assumed to consist of independent categorical groups. In our case, this refers to the 4 platforms - Tumblr, Reddit, Twitch, and 4chan, which are independent since each thread belongs to only one of the platforms. Finally, the observations should be independent. In our case, since each comment only contributes to one thread, the threads can be said to be independent of each other. The test works by first ranking all the threads for each feature. Then, the test statistic is computed according to formula 2

$$H_f = (N-1) \frac{\sum_{p \in platforms} n_p (\bar{r_{fi}} - \bar{r_f})^2}{\sum_{p \in platforms} \sum_{i=1}^{n_p} (r_{pfi} - \bar{r_f})^2} \tag{2}$$

where N = 405 is the total number of threads, n_p is the number of threads from platform p, r_{pfi} is the rank of thread i from platform p in terms of feature f, $\bar{r_{pf}}$ is the average rank of all threads from platform p in terms of feature f, and $\bar{r_f}$ is the average rank of feature f. The null hypothesis of the test is that the population means do not differ from one another. A sufficiently small p-value leads to the rejection of the null hypothesis and the conclusion that at least two platforms differ in terms of the feature being tested. If that is the case, a post hoc pairwise test must be applied to determine which pairs of means are significantly different.

We use a Conover-Iman test (Conover and Iman, 1979), available through the scikit-posthocs library, as a post hoc test to see which platforms differ in terms of each feature. This is a non-parametric test based on T values which does not assume a normal distribution and is meant to follow the rejection of a Kruskal-Wallis null hypothesis. The null hypothesis for each pair of groups is that there is no difference between them. A sufficiently small p-value leads to the rejection of this hypothesis and to the conclusion that the difference in group means is statistically significant. These results will help answer our first two research questions, determining whether different platforms differ significantly in terms of community and politeness.

4.3.3 Regresion

Finally, we are performing a linear regression using the aggregate politeness feature as a dependent variable, with community and each of the platforms being independent variables. The model used is given in equation 3.

$$\begin{aligned} y_i = &\beta_0 + \beta_1 x_{i1} + \beta_2 \mathbb{1}(x_{i2} = 2) + \beta_3 \mathbb{1}(x_{i2} = 3) + \beta_4 \mathbb{1}(x_{i2} = 4) \\ &+ \beta + 5x_{i1}\mathbb{1}(x_{i2} = 2) + \beta + 6x_{i1}\mathbb{1}(x_{i2} = 3) + \beta + 7x_{i1}\mathbb{1}(x_{i2} = 4) \end{aligned} \tag{3}$$

In this model, y_i is the dependent variable "Politeness" value of thread i, x_{i1} is the independent variable "Community" value of thread i and x_{i2} is the independent variable "Platform" corresponding to thread i where Tumblr = 1, Reddit = 2, and 4chan = 3, with this variable only being used inside identity functions. Linear regression assumes a linear relationship between the dependent variable and the independent variable. If the relationship exists but is not linear, this is a potential limitation of our study. The observation must be independent, which we have already argued must be the case since each comment belongs to only one thread and each thread belongs to only one platform. There is also an assumption that the errors are normally distributed. This will be checked in the Results section by using a Q-Q plot to check if the distribution of the residuals resembles a normal distribution. The results will help answer our third research question, determining if community has an impact on politeness and whether this impact differs across platforms.

5 Results

5.1 Feature Analysis

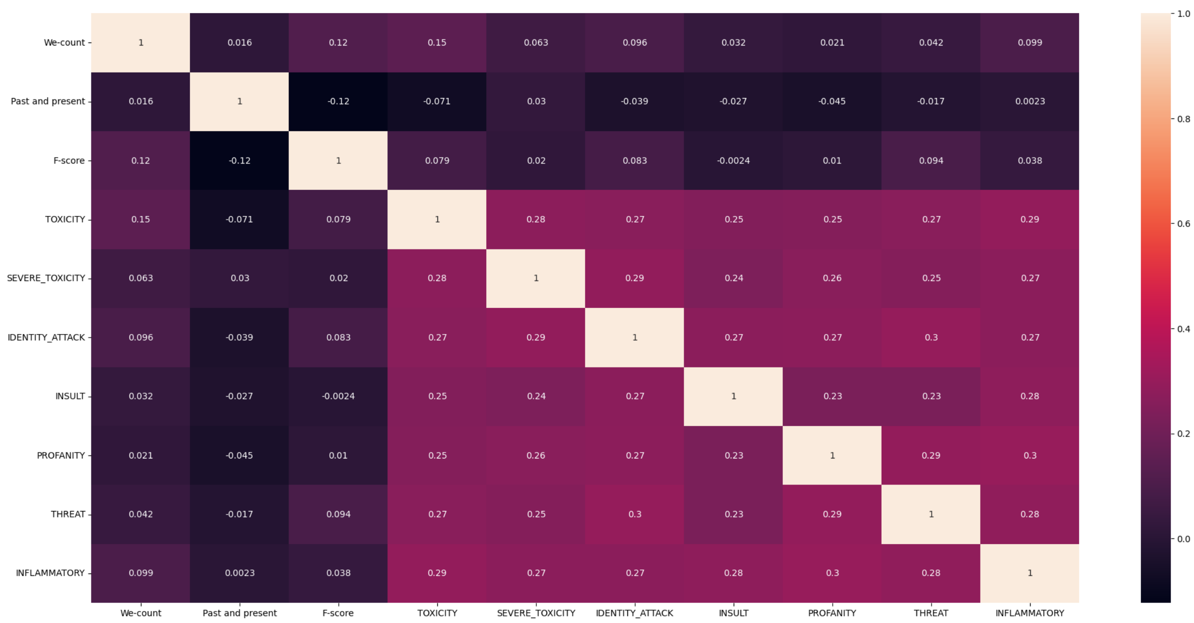


Figure 6: Correlation table between individual features

Figure 6 shows the correlations between the individual features. None of the individual pairs is strongly correlated, which gives validity to our regression analysis.

Table 2 shows the descriptive statistics of the community feature. Tumblr and Reddit tend to score the highest on this metric, with Reddit having by far the highest value encountered, both of them having positive means and similar standard deviations. 4chan scores very close to 0 on average, which represents the mean over all 4 platforms, with a slightly lower standard deviation. Twitch scores the lowest, with a negative mean, a minimum slightly lower than 4chan, but a very small maximum compared to the other 3 platforms. 4chan and Twitch also have similar standard deviations. The most surprising aspect is that of Twitch, which directly contradicts our hypothesis that Twitch must score high on community. Due to its heavy use of stickers and inside jokes that don't tend to spread very quickly outside the Twitch user base, we expected Twitch to have q high level of contextuality, and we expected that the userbase forming a collective entity might be reflected in the we-count. Appendix A shows the normalised values of these values, and the results show that the F score is indeed the highest among all three, but the we-count is average and the emotion and verb tense scores are both lower than all other platforms. On the other hand, emotion and verb tense seem to be why Reddit has the highest community scores.

Table 3 Shows the same results for the aggregate politeness feature. Reddit once again has the highest features, Twitch is almost exactly equal to 0 on average, and Tumblr is in-between them. 4chan has very low scores compared to the other three, with a strongly negative mean that seems to be driven less by small highs and instead by a very small minimum value. This is unsurprising given 4chan's "toxic" reputation.

Platform	Min	Mean	Max	StDev
Tumblr	-1.85	0.34	3.28	1.07
Reddit	-1.91	0.45	6.39	1.05
Twitch	-2.62	-0.80	0.66	0.63
4chan	-2.29	0.04	2.07	0.62

Table 2. *Community*

Platform	Min	Mean	Max	StDev
Tumblr	-2.29	0.22	1.53	0.92
Reddit	-1.31	0.62	1.66	0.74
Twitch	-1.49	0.03	1.09	0.60
4chan	-3.11	-0.87	1.18	1.04

Table 3. *Politeness*

5.2 Tests

The results of the Shapiro-Wilk test lead to the rejection of the normality hypothesis in half of the cases for Community and Politeness and in most cases for individual features. Thus, we will only use tests which do not rely on the assumption that the variables follow a normal distribution. The Kruskal-Wallis test statistics for Community and Politeness are 123.95, and 111.68 respectively, with p values equal to $1.08 * 10^{-26}$ and $4.75 * 10-24$ which leads to the rejection of the hypothesis that the platforms do not differ in terms of these features. In fact, the full result shown in Appendix B shows that the null hypothesis is strongly rejected for all features other than verb tense.

Table 4 shows the results of the pairwise Conover-Iman test. Tumblr-Reddit and Tumblr-4chan differ with a significance level of 10%, and every other combination differs with less than 1% significance. Table 5 shows that all combinations differ at a 1% level.

	tumblr	reddit	twitch	4chan
tumblr	1			
reddit	0.084	1		
twitch	0.000	0.000	1	
4chan	0.067	0.000	0.000	1

Table 4. *Conover-Iman test on Community*

	tumblr	reddit	twitch	4chan
tumblr	1			
reddit	0.000	1		
twitch	0.013	0.000	1	
4chan	0.000	0.000	0.000	1

Table 5. *Conover-Iman test on Politeness*

The results of the Conover-Iman text, combined with the earlier descriptive statistics show that the differences observed are not due to chance or random variance in the samples gathered but instead are statistically significant enough to show a difference between the way people talk to each other on all these 4 websites.

5.3 Regression

	coef	std err	$P > \|t\|$
const	0.6685	0.091	0.000
x1	-0.1145	0.080	0.152
x2	-0.4183	0.127	0.001
x3	-0.6623	0.161	0.000
x4	-1.5484	0.124	0.000
x5	0.0230	0.112	0.837
x6	0.0886	0.153	0.562
x7	0.3069	0.157	0.052

Table 6. *Regression results*

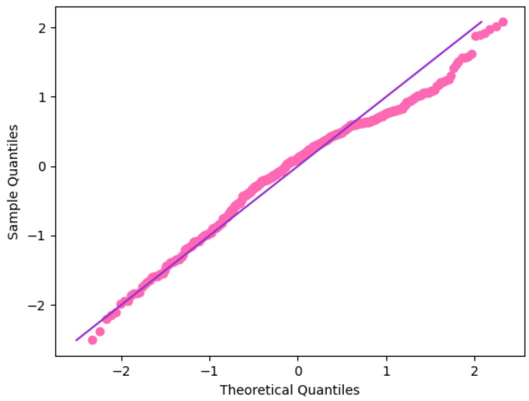


Figure 7: A Q-Q plot of the regression residuals

Table 6 shows the results of the regression analysis. Reddit is taken as a benchmark. The constant represents the average expected politeness score of a Reddit thread with a globally average level of community signals. X1 represents the difference between this globally average thread and the actual community score of a Reddit thread. Thus, a Reddit thread with a community score of C is expected to have a politeness score of P = 0.67 - 0.11C. Thus, for Reddit threads, it holds that the higher the community score, the lower the politeness, on average. The significance level of this, however, is 15%, meaning that the data suggests a 15% chance to obtain these results due to chance, even if community does not, in fact, have any effect on politeness. X2 represents the difference between the expected politeness score of a globally average Reddit thread and a globally average Tumblr thread. The significance level is 1%, suggesting strong evidence that Tumblr and Reddit differ in terms of average levels of politeness, which is confirmed by previous tests. Specifically, the expected score of a Tumblr thread with a globally average community score is 0.67-0.42=0.25, which is in line with what previous results suggested. X3 and X4 represent the differences for twitch and 4chan, respectively. Both of them have confidence levels of over 99%, and both of them show that Twitch and 4chan have significantly lower average politeness scores, with 0.01 and -0.88, respectively. X5 represents the difference between the effect of community on a Reddit thread and the effect of community on a Tumblr thread. Specifically, the result suggests that for a community level of C, the expected politeness level is P = 0.25 - 0.9C. This would suggest that on Tumblr, while the effect of community on politeness is still negative, it is not as strong as for Reddit. However, not only is this difference very small, but the significance level of 84% suggests a strong possibility that the result is given by random variance in the data, and the effect of community on politeness is the same on both platforms. X6 represents the same difference as X5, but this time between Reddit and Twitch. The coefficients suggest that the expected politeness score of a Twitch chat with community C is P = 0.01 - 0.03. This is still a negative effect, but much less strong than that of Reddit. On the other hand, the confidence level around 50% means we cannot rely on this result as the chance it could be due to chance is over 50%. Finally, X7 represents the difference between the average effect of community on 4chan threads compared to Reddit threads. The results show an expected level of politeness of P = -0.88 + 0.20C. This time, the increase in community is not only positive but has a significance level of 5%, suggesting that the difference observed is very likely due to differences in the actual effects.

Figure 7 shows the results of generating a Q-Q plot from the residuals. A linear shape at a 45°angle suggests that the points are normally distributed. Thus, the results of the regression can be taken to be valid as the residual normality assumption is plausible.

6 Dicussion and Conclusion

There has not yet been a consensus on the effects of anonymity on politeness or community. Some evidence suggests that anonymous users are more likely to show signs of toxicity, but the number of platforms and boards sampled is usually limited. Furthermore, research has hypothesised that a stronger sense of community and adherence to group norms might actually lead to a reduction in politeness in some cases, but this has not been tested on large samples.

We looked at four social media platforms - Tumblr and Reddit, with little user anonymity, Twitch, where users have persistent accounts but the "individual" usually becomes part of the collective "chat", and messages cycle quickly and disappear just as quickly, and 4chan, where users have full anonymity and messages disappear from the platform after a few days.

Our results suggest that the websites with lower user anonymity (Tumblr and Reddit) score higher both on signals of community and on signals of politeness. While this was expected, the fact that there is a significant difference between the two was very unexpected. Since Tumblr and Reddit have the same type of user identity, the differences between the two cannot be explained by anonymity, and instead, there must be unobserved underlying factors that affect the results. Between Twitch and 4chan, Twitch ends up scoring very low on community, while 4chan scores are very low on politeness. If we were to attribute these results to anonymity, one possible explanation would be that the ephemerality of a Twitch message and the "chat" identity causes the individual "I" to disappear, with users scoring low on trust and sociability as they do not have the opportunity to open up to the community as an individual. What is surprising is that the collective "we" also disappears on Twitch, having low maxima of group stability and we-count, which is what ends up driving the low averages. An alternative explanation for low community markers lies in the use of the platform - with most users watching the streamer's gameplay and most messages referring to what they are watching instead of references to individual users or to the chat as a collective. This suggests a potential limitation of our study in the way the idea of community is assessed and computed, as it becomes hard to tell whether the lack of community signals is indeed due to a lack of a strong community identity or due to the nature of the stream-watching activity. On the other hand, the low levels of politeness on 4chan are very much in line with our hypotheses, previous research, and the general reputation of the platform.

When looking at the effect of community on politeness, we hypothesised that higher scores of community markers might, on occasion, lead to lower scores of politeness. Furthermore, we expected that the exact relationship might differ from platform to platform, with the final research question being how this effect differs with anonymity. The former hypothesis was confirmed, with three of the four platforms studied showing a negative relationship between community and politeness. The small and insignificant difference between Tumblr and Reddit conforms to our expectations. Twitch also does not differ significantly from Reddit, although the confidence level is still higher than that of Tumblr vs Reddit. Thus, the idea that a low sense of individuality and high message turnover affects the community norms regarding politeness cannot be confirmed but remains an interesting possibility to study. The most unexpected result of the study is the strongly positive relation between community markers and politeness of 4chan. Since the platform has such a reputation for tolerating and welcoming toxic behaviour, one would be confident in the assumption that toxic behaviour or "trolling" would be not only tolerated but encouraged on 4chan. Our results, however, suggest the exact opposite.

Overall, our results show that of the two platforms that were deemed to "score" higher in anonymity, the platform with user impermanence ends up being antisocial - answering our first research question, while the one with message impermanence ends up being asocial - answering our second research question.

The significant differences between the two non-anonymous platforms suggest unobserved factors affecting the results and suggest that any interpretation of the results being driven by anonymity must be assessed critically, in spite of our attempts to isolate the sub-community and topic of conversation to that of video games.

One of the most interesting and valuable contributions of our study is the confirmation that a stronger sense of community might lead to higher levels of impoliteness, which answers our third research question, but the most unexpected finding is that this is not the case on 4chan, a platform notorious for high levels of toxicity. If we are to accept our assumption that the only difference between platforms is anonymity, this suggests that on 4chan, it is anonymity alone that drives toxicity, not through but in spite of the tacit community expectations.

Our results seem to open as many questions about the topic as we manage to answer. This proves to be an interesting direction for future research to better understand and isolate the underlying mechanism of the effect of anonymity on politeness and community. The largest limitation of this paper is the questionable assumption that anonymity is the only difference at play between the platforms investigated. The differences between Reddit and Tumblr can serve as a sanity check - only when this difference is no longer statistically significant can the assumption of anonymity driving the results be reliably claimed. Another limitation of our study, which can be addressed in future research, is the assumption of a linear relationship between community and politeness. It could be argued that while low levels of politeness might stem from a familiar forgoing of traditional decorum - profanity or threats used lightly and jokingly, very low levels represent high levels of toxicity - profanity and threats used to scare and intimidate. Thus, it might be the case that while low-level impoliteness might be encouraged by a higher sense of community, high-level toxicity might not, leading to a nonlinear relation between the two. Finally, we believe there could be value in expanding the types of platforms considered. While Tumblr and Reddit have persistent user identities, these identities don't usually match real-life identities, with people relying at least somewhat on the aspect of anonymity on these websites. Having access to interactions carried over platforms that require users to use real names and photos - such as Facebook - can provide a new interesting level to this area of research.

References

boilingpenguin (2018, August). 4chan-scraper. `https://github.com/boilingpenguin/4chan-scraper.git`.

Code, J. and N. Zap (2009, 01). Social identities, group formation, and the analysis of online communities. *Handbook of Research on Social Software and Developing Community Ontologies*, 86–101.

Connolly, T., L. Jessup, and J. Valacich (1990, 06). Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science 36*, 689–.

Conover, W. J. and R. L. Iman (1979, 2). Multiple-comparisons procedures. informal report.

Dillon, K. P., R. L. Neo, and N. Seely (2015, Oct.). Civil keystrokes: Examining anonymity, politeness, and civility in online newspaper forums. *AoIR Selected Papers of Internet Research 5*.

Durant, L. E., M. P. Carey, and K. E. E. Schroder (2002, October). Effects of anonymity, gender, and erotophilia on the quality of data obtained from self-reports of socially sensitive behaviors. *J. Behav. Med. 25*(5), 438–467.

Khalid, O. and P. Srinivasan (2020, May). Style matters! investigating linguistic style in online communities. *Proceedings of the International AAAI Conference on Web and Social Media 14*(1), 360–369.

Kim, J. (2019, Aug). Twitch.tv chat log data.

Kruskal, W. H. and W. A. Wallis (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47*(260), 583–621.

Laerd Statistics (2018). Kruskal-wallis h test using spss statistics. https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php.

Lapidot-Lefler, N. and A. Barak (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior 28*(2), 434–443.

Locher, M. A. and R. J. Watts (2008). *Chapter 4. Relational work and impoliteness: Negotiating norms of linguistic behaviour*, pp. 77–100. De Gruyter Mouton.

McEwan, B. (2016). Communication of communities: linguistic signals of online groups. *Information, Communication & Society 19*(9), 1233–1249.

Papasavva, A., S. Zannettou, E. De Cristofaro, G. Stringhini, and J. Blackburn (2020). Raiders of the lost kek: 3.5 years of augmented 4chan posts from the politically incorrect board. *Proceedings of the International AAAI Conference on Web and Social Media 14*, 885–894.

pistocop (2022, February). subreddit-comments-dl. `https://github.com/pistocop/subreddit-comments-dl.git`.

Postmes, T., R. Spears, K. Sakhel, and D. Groot (2001, 10). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin 27*, 1242–1254.

Shapiro, S. S. and M. B. Wilk (1965). An analysis of variance test for normality (complete samples). *Biometrika 52*(3/4), 591–611.

Smith, J., D. Terry, and M. Hogg (2007, 03). Social identity and the attitude-behaviour relationship: Effects of anonymity and accountability. *European Journal of Social Psychology 37*.

Vogel, C. (2014). Anonymous FreeSpeech. In K. Agmad and C. Vogel (Eds.), *DIMPLE: DIsaster Management and Principled Large-scale information Extraction for and post emergency Logistics*, pp. 26–31. The 9th edition of the Language Resources and Evaluation Conference. Reykjavik, Iceland, May 26-31, 2014.

Young, R., S. Miles, and S. Alhabash (2018). Attacks by anons: A content analysis of aggressive posts, victim responses, and bystander interventions on a social media site. *Social Media + Society 4*(1), 2056305118762444.

Zhang, K. and R. Kizilcec (2014, 05). Anonymity in social media: Effects of content controversiality and social endorsement on sharing behavior. *Proceedings of the International AAAI Conference on Web and Social Media 8*, 643–646.

7 Appendix A: Individual features descriptive statistics

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.13	0.54	0.13
Reddit	0.00	0.10	1.00	0.14
Twitch	0.01	0.08	0.15	0.02
4chan	0.00	0.11	0.67	0.10

Table 7. *Emotion*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.06	0.72	0.14
Reddit	0.00	0.03	0.38	0.07
Twitch	0.00	0.04	0.22	0.04
4chan	0.00	0.03	1.00	0.10

Table 8. *We-count*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.22	0.73	0.17
Reddit	0.00	0.26	1.00	0.18
Twitch	0.04	0.20	0.50	0.07
4chan	0.00	0.22	0.62	0.12

Table 9. *Verb tense*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.37	0.78	0.15
Reddit	0.10	0.33	0.63	0.11
Twitch	0.40	0.62	1.00	0.15
4chan	0.20	0.41	0.74	0.08

Table 10. *F score*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.23	0.93	0.25
Reddit	0.00	0.18	0.86	0.22
Twitch	0.01	0.29	0.88	0.22
4chan	0.01	0.40	1.00	0.29

Table 11. *Toxicity*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.22	0.76	0.21
Reddit	0.00	0.13	0.75	0.18
Twitch	0.01	0.22	0.70	0.20
4chan	0.01	0.40	1.00	0.27

Table 12. *Severe toxicity*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.21	0.84	0.22
Reddit	0.00	0.15	0.96	0.22
Twitch	0.01	0.27	0.71	0.20
4chan	0.00	0.41	1.00	0.27

Table 13. *Identity attack*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.25	0.83	0.24
Reddit	0.00	0.16	0.72	0.19
Twitch	0.01	0.23	0.77	0.19
4chan	0.01	0.40	1.00	0.28

Table 14. *Insult*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.25	0.78	0.25
Reddit	0.00	0.16	0.83	0.22
Twitch	0.01	0.23	0.93	0.21
4chan	0.01	0.41	1.00	0.29

Table 15. *Profanity*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.20	0.97	0.21
Reddit	0.00	0.15	0.96	0.21
Twitch	0.01	0.27	0.82	0.21
4chan	0.01	0.41	1.00	0.30

Table 16. *Threat*

Platform	Min	Mean	Max	StDev
Tumblr	0.00	0.22	0.90	0.21
Reddit	0.00	0.16	0.75	0.21
Twitch	0.00	0.26	0.69	0.18
4chan	0.01	0.41	1.00	0.28

Table 17. *Inflammatory*

Platform	Min	Mean	Max	StDev
Tumblr	0.14	0.57	0.85	0.14
Reddit	0.25	0.55	1.00	0.12
Twitch	0.19	0.50	0.90	0.09
4chan	0.00	0.49	0.90	0.15

Table 18. *Politeness*

8 Appendix B: Kurskal - Wallis test results of the features

Feature	Statistic	p-value
Emotion	15.48	0.001446
We - count	58.45	1.26E-12
Verb tense	10.11	0.017632
F score	183.69	1.41E-39
Toxicity	46.09	5.41E-10
Severe Toxicity	63.6	9.96E-14
Identity Attack	69.34	5.89E-15
Insult	42.9	2.58E-09
Profanity	58.41	1.28E-12
Threat	63.36	1.12E-13
Inflammatory	59.45	7.68E-13
Politeness	125.93	4.07E-27

Table 19. *Kruskal-Wallis test*

Contribution

- Teona Banu - Chair: Detailed the research questions, hypotheses, and contributions to literature. Contributions to the literature review on the topic of community. Wrote the Data section description as well as gathered and described the Tumblr and 4chan data sets, and described the Twitch data set. Performed and described the process of the extraction of the we-count, verb tense, F-score, (severe) toxicity, identity attack, insult, profanity, threat, and inflammatory speech features. Computed aggregate features. Normalised all the features. Performed the Shapiro-Wilk test on all the features. Performed the Kurskal Wallis test and Dunn post hoc test on all the features. Performed and described the methodology of the regression analysis. Wrote the results section. Wrote the discussion and conclusion section.
- Vivek Bhadula - Ambassador: Contributed to the literature section; involved finding and summarizing two relevant papers that were related to the topic being studied. This involved searching through academic databases and identifying articles that were most relevant to the research question. Once these articles were identified, I carefully read and analyzed them, highlighting the key findings and ideas that were presented. Formatted the entire literature review. I also generated correlation tables that helped to illustrate the relationships between different variables that were being studied. These tables were carefully constructed and analyzed to ensure that they accurately reflected the data that had been collected and were presented in a clear and easy-to-understand format.
- Shuncheng Cai - Recorder: Contributed to the Literature section, specifically to writing the overview of all the relevant literature found. searched the relevant literature and initially mapped out the available algorithms and implemented models for the project. Performed and described the methodology of the data preprocessing, specifically data cleaning. Performed and described the methodology of extracting the politeness feature.
- Xiaohan Shi - Accountant: Clean up data sets, contribute to crawling data on reddit, and contribute to literature reviews on 4chan topics. Write codes to analyze the negative emotion and positive emotion of 4 websites and obtain the results, and standardize the results. Summarize all the initial values for the 13 features and calculate their mean, maximum, minimum, standard deviation and median. Write the Introduction, key words of the paper and the Negative or Positive Emotions part of the Methodology.
- Mengqi Wang - Verifier: Write code to crawl the Reddit dataset, clean and format the data. Contributed in analyzing negative and positive sentiment from 4 sites, obtaining results, and normalizing the results. Created descriptive statistics tables on the dataset, counting the minimum, maximum, mean, median, and mean standard deviation of the number of comments in each thread for each platform and the number of words per comment in each thread. Write the data description in the reddit section, contributing to the literature review section and the sentiment analysis section.

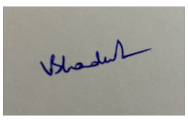
Teona Banu	Vivek Bhadula	Shuncheng Cai	Xiaohan Shi	Mengqi Wang
Banu Teona		SHUNCHENG CAI	Xiaohan Shi	Mengqi Wang

Table 20. *Signature*

The meeting agendas are available at: Meeting Agendas