

A Longitudinal Measurement Study of 4chan’s Politically Incorrect Forum and its Effect on the Web

Gabriel Emile Hine[†], Jeremiah Onaolapo[‡], Emiliano De Cristofaro[‡], Nicolas Kourtellis^{*}, Ilias Leontiadis^{*}, Riginos Samaras[°], Gianluca Stringhini[‡], Jeremy Blackburn^{*}

[†]Roma Tre University [‡]University College London

^{*}Telefonica Research [°]Cyprus University of Technology

ABSTRACT

Although it has been a part of the dark underbelly of the Internet since its inception, recent events have brought the discussion board site 4chan to the forefront of the world’s collective mind. In particular, /pol/, 4chan’s “Politically Incorrect” board has become a central figure in the outlandish 2016 Presidential election. Even though 4chan has long been viewed as the “final boss of the Internet,” it remains relatively unstudied in the academic literature.

In this paper we analyze /pol/ along several axes using a dataset of over 8M posts. We first perform a general characterization that reveals how active posters are, as well as how some unique features of 4chan affect the flow of discussion. We then analyze the content posted to /pol/ with a focus on determining topics of interest and types of media shared, as well as the usage of hate speech and differences in poster demographics. We additionally provide quantitative evidence of /pol/’s collective attacks on other social media platforms. We perform a quantitative case study of /pol/’s attempt to poison anti-trolling machine learning technology by altering the language of hate on social media. Then, via analysis of comments from the 10s of thousands of YouTube videos linked on /pol/, we provide a mechanism for detecting attacks from /pol/ threads on 3rd party social media services.

0. EXECUTIVE SUMMARY

The web has a lot of dark corners, and since 2003, 4chan.org has been considered one of the darkest. Known for memes, trolling, and more, 4chan is an anonymous bulletin board system. 4chan has most recently come under public scrutiny with narrative pushed by its politically incorrect board, /pol/, with respect to the 2016 US Presidential election. Although it is a bit absurd, /pol/ has, somehow, managed to place itself at the center of world politics. Considering its clear impact on society, 4chan in general has been relatively unexplored. In this paper, we begin to rectify this gap in knowledge by perform a longitudinal study of /pol/. Using a dataset of over 8 million posts crawled since June 20, 2016, we study /pol/ along several axes.

We begin by understanding posting behavior on /pol/ and two other 4chan boards: /sp/ (“sports”) and /int/ (“international”). We find that 4chan’s unique “bump” system seems to be quite successful at ensuring fresh content is available across all three boards. We also find differences in moderation across each board, as well as the

use of 4chan’s tripcode system that allows users to have a persistent identity in the presence of 4chan’s default “Anonymous” posting behavior. Using a special feature on /pol/, we also gain some insight into the number of *unique* users that participate in threads, finding that while /pol/ is anonymous, it is also full of many voices. Finally, we look at posting behavior of different countries, finding evidence that while American’s dominate the conversation on /pol/, many countries are well represented in terms of their Internet using population.

Next we perform content analysis on /pol/. We find that, by far, the most popular links shared are to YouTube, and that “main stream” news sites are less popular than “alt-right” leaning sources. When looking at the images posted on /pol/, we find that it is almost entirely “original content:” 70% of unique images are posted only once and 95% are posted less than 5 times. We then delve into the type of text that /pol/ posts by looking at the most “popular” hate speech used and showing varying levels of hate speech usage by different countries.

Finally, we look into “raiding” behavior on /pol/. Raids are similar to DDoS attacks except that instead of trying to interrupt the service on a network level, they attempt to disrupt the *community* that calls the service home, e.g., by harassing users. We first perform a quantitative case study of /pol/’s recent “Operation Google” which was an attempt to replace hate words with more innocuous terms (in particular, names of Internet companies). We show that Operation Google had a substantial impact on /pol/ and is still somewhat in effect. However we find little evidence of direct impact on Twitter: it seems that Operation Google was mostly a success in terms of propaganda via media coverage. Finally, we explore raiding behavior on YouTube comments. We show that there is statistically significant evidence that certain YouTube comments linked to by /pol/ experience a peak in activity. Then, using cross-correlation we estimate the synchronization lag between posts on /pol/ and YouTube comments. Finally, we show that as the synchronization lag approaches zero, we see an increase in the hate words that appear in the YouTube comments. I.e., we find statistically significant evidence that /pol/ is attacking YouTube comments.

1. INTRODUCTION

Over the past few years, the web has evolved from providing the means to communicate and exchange information, to playing a key role in several aspects of our society. For instance, today, we use the Internet for entertainment, work, politics, social interactions, finding romantic relationships, and so on. Moreover, it is also a source for new culture—whether this is considered good or bad. At the same time, the web has fed into new socio-technical concerns,

Corresponding author: jeremyb@tid.es

ranging from crime [3], to privacy [19], to “net overload” [25].

Among the most worrying threats, harassment and hate speech have become increasingly prevalent [7, 14]. The web’s global communication capabilities, as well as a number of platforms built on top of them, often enable previously isolated, and possibly ostracized, members of fringe political groups and ideologies to gather, converse, organize, execute, and spread their culture of hate [23].

In particular, 4chan.org has emerged as one of the most impactful generators of online culture. Launched in 2003 by Christopher Poole (at the time identifying himself with the pseudonym moot), and acquired by Hiroyuki Nishimura in 2015, 4chan is an imageboard site, built around a typical bulletin-board model where users can create posts and others can reply in kind. On 4chan, an “original poster” (OP) creates a new thread by making a post, with one single image attached, to a particular board with a particular interest focus. Other users can reply, with or without images, and possibly add references to previous posts, quote text, etc. One of 4chan’s most important aspects is its anonymous nature: there is no login-based account system, and the overwhelming majority of posts/replies are featured as authored by “Anonymous” [6].¹

Not only have significant chunks of “Internet culture” and memes² arisen from 4chan, but so have political movements like “Anonymous” and positive actions such as catching animal abusers [1]. At the same time, 4chan is also one of the darkest corners on the Internet, featuring porn, hate speech, trolling, and even murder confessions by users [12], as well as a platform to coordinate distributed denial of service attacks [2].

Despite its influence and coverage in the media³, 4chan remains largely unstudied from a scientific perspective. In this paper, we start addressing this gap by focusing on one sub-community, namely, “/pol/”, i.e., the “Politically Incorrect” board. /pol/ is, to some extent, considered a “containment” board, allowing users to discuss generally distasteful content (even by 4chan standards) without disturbing the operations of other boards. Even though /pol/’s contents do revolve around politically incorrectness, a simple visual scan of discussions at any given time makes it clear that the majority of posters subscribe to the “alt-right” movement, exhibiting characteristics of xenophobia, social conservatism, racism, and, generally speaking, hate.

Overview & Contributions. This paper presents a multi-faceted analysis of /pol/, using a dataset of 8M posts from over 216K conversation threads that we have collected over a 2.5-month period (Section 2). We perform a general first-of-its-kind characterization of /pol/, focusing on overall posting behavior, as well as exploring how the intricacies of 4chan’s system influence the way discussions proceed (Section 3). Next, we explore the types of content that /pol/ shares, including 3rd party links, images, and the use of hate speech (Section 4). Finally, we show that /pol/’s hate-filled vitriol is *not* contained within /pol/, or even 4chan, and in fact has substantive effects on conversations taking place on other computer mediated communication platforms via a phenomenon called “raids” (Section 5). We provide a quantitative case study of /pol/’s attempt to poison anti-trolling machine learning technology by altering the language of hate on social media. Then, via analysis of comments from the 10s of thousands of YouTube videos linked on /pol/, we provide a mechanism for detecting attacks from /pol/ threads on 3rd party social media services.

¹Note that 4chan employs a mechanism called “tripcodes” which can act as a proxy for a user name – see Section 3.2.

²For readers unfamiliar with memes, we suggest a review of the documentary available at <https://www.youtube.com/watch?v=dQw4w9WgXcQ>.

³At the time of this writing, speculation on 4chan’s financial solvency is making headlines, e.g., <http://www.bbc.com/news/technology-37563647>.



Figure 1: Examples of typical /pol/ threads. Thread A illustrates the derogatory use of “cuck” in response to an image of Bernie Sanders. Thread B shows a casual call for genocide with an image of a woman’s cleavage as well as a “humorous” response. Thread C illustrates /pol/’s fears that a possible withdrawal of Hillary Clinton due to health issues would guarantee Donald Trump’s loss. Thread D is dedicated to “Kek,” the god of memes via which /pol/ “believes” they influence reality.

2. PRELIMINARIES

2.1 4chan

4chan is an imageboard site, similar to a typical bulletin-board site, although it actually has several unique characteristics, which we now review. On 4chan, an “original poster” (OP) creates a new thread by making a post, with one single image attached, to a particular board with a particular interest focus. Other users can post in that thread, with or without images⁴, and possibly add references to previous posts in the thread by replying to or quoting portions of a post.

Boards. 4chan separates conversation into different areas of interests known as “boards.” At the time of this writing, 4chan has 69 boards split into 7 high level categories ranging from “Japanese Culture” (9 boards) to “Adult (NSFW)” (i.e., porn, 13 boards).

In this paper, we focus on /pol/, the “politically incorrect” board. The rules of /pol/ are relatively simple with threads getting deleted pretty much only if considered off-topic.⁵

Figure 1 shows four typical /pol/ threads. Besides the content, the figure also illustrates 4chan’s “reply” feature (“>12345” indicates a reply to post “12345”), the flag system indicating the posters location, the prevalence of the default “Anonymous” poster name, and poster IDs (the colored hash text next to the poster’s name), each of which we will explain in more detail a little later.

We also compare /pol/ to the behavior on two other boards: sports (/sp/) and international (/int/). The former focuses on sports and

⁴Note, to the best of our knowledge, 4chan only allows uploading of images that are *not* already posted in a live thread within a given board.

⁵<http://boards.4chan.org/pol/>

athletics, while the latter on different cultures, languages, etc. Both /sp/ and /int/ are considered “safe for work” boards, and are, in theory, more heavily moderated.

Anonymity. On 4chan, users do not need to register an account to participate in the community. Anonymity is the default (and preferred) behavior, although 4chan does support *some* degree of permanence and identifiability for users. Specifically, while the default username is “Anonymous”, users are allowed to enter a name along with their posts: since there is no account system, anyone is free to use whatever name they wish, so usernames do not provide identity. However, “*tripcodes*”, i.e., hashes of user-supplied passwords, can be used to link threads from the same user across time, and providing a way for users to verify their (pseudo-)identity to others. Note that tripcodes are generally considered somewhat “un-cool” and require additional effort from the user [6].

Also, on some boards (including /pol/), intra-thread trolling led 4chan to introduce “poster IDs.” Within a given thread (but *only* that thread), each poster is given a unique ID that appears along with their post. This countermeasure preserves the overall “Anonymous” theme, but mitigates low-effort sock puppeteering within a thread. To the best of our knowledge, poster IDs are determined via a combination of cookies and IP based client identification.

Ephemerality. 4chan threads are pruned after a relatively short period of time, using a “*bumping*” system. When users visit a board, threads from that board’s catalog are presented, with threads having the most recent post appearing first. The number of threads in the catalog is limited, so creating a new thread results in the one with the least recent post being removed. Although this ensures that older content is removed, it does not prevent certain threads from dominating the board. For instance, users wishing to disrupt the board could simply bump a thread, e.g., every second, significantly increasing its chance to remain in the catalog indefinitely.

To address this potential issue, 4chan implements so-called bump and image limits – i.e., after a thread is bumped N times or has M images posted to it (with N and M being board-dependent), new posts to it will no longer bump it up. Therefore, while the thread can still receive new posts, it will eventually be purged as new threads are created. Originally, when a thread fell out of the catalog, it was permanently gone, however, 4chan has recently implemented an archive system for a subset of boards. That is, once a thread is purged, no new posts are possible but its final state is archived for some period of time (7 days at the time of this writing).

Sticky Threads. Alongside regular threads, 4chan also features so-called sticky threads. These follow special rules: they are always “stuck” at the top of the catalog, do not have any thread limits, and are configured to keep the most recent 1,000 posts. Sticky threads in /pol/ are often created in response to special events, e.g., the 2016 Republican National Convention.

Flags. Certain boards on 4chan (including /pol/, /sp/, and /int/) additionally include (along with each post) the flag of the country the user posted from, based on IP geo-location. This somewhat reduces the ability for users to “troll” each other by e.g., claiming to be from a country where some event is happening, although geo-location can be fooled using VPNs and proxies.

Moderation. 4chan does have a kind of moderation system, involving so-called “janitors,” i.e., volunteers recruited every once in a while from the user base⁶. Janitors are given limited tools that allow them to prune posts and threads, as well as recommend bans to more “senior” 4chan employees. While the internal workings

⁶<https://www.4chan.org/janitorapp>

	/pol/	/sp/	/int/	Total
Threads	216,783	14,402	24,873	256,058
Posts	8,284,823	1,189,736	1,418,566	10,893,125

Table 1: Number of threads and posts crawled for each board.

of 4chan as a legal entity are beyond our knowledge, we are under the impression it is a relatively small operation that is constantly fighting to stay solvent.⁷ Thus, while generally speaking, janitors are not well respected by 4chan users and are often mocked for their perceived love for the modicum of power they have⁸, they do contribute to 4chan’s continuing operation.

2.2 Datasets

We began crawling 4chan on June 30, 2016 using their JSON API⁹, retrieving /pol/’s thread catalog every 5 minutes. Our crawler compares the threads that are currently live to those in the previously obtained catalog, then, for each thread that has died, we retrieve a full copy of it from 4chan’s archive, which allows us to obtain the full/final contents of a thread.

For each post in a thread, the 4chan API gives us, among other things, the post’s number, its author (almost always “Anonymous”), the timestamp the post was made, the contents of the post¹⁰, etc. Although our crawler does not save images, the 4chan API also includes some meta data for images that were posted, e.g., the name the image was uploaded with, the dimensions (width and height) of the image, file size, and an MD5 hash of the image.

Since only the most recent 1,000 posts are kept in sticky threads, our crawl-after-death would fail to retrieve all posts, therefore, we crawl every sticky thread in the catalog every 5 minutes, which, in our experiments, is frequent enough to capture all activity.

On August 6, 2016 we began crawling /sp/, 4chan’s sports boards, and on August 10, 2016 we began crawling /int/, the international board. Table 1 provides a high level overview of our dataset. We note that for about 6% of the threads, our crawler gets a 404 error when retrieving them from the archive: from a manual inspection, it seems that this is due to “janitors” (i.e., moderators) removing threads for violating rules.

In addition to 4chan, we also collected YouTube comments for videos that were linked on 4chan and data from Twitter; see Section 5 for details. While our crawler is continuously running, for all analysis in this paper *except* Section 5 we only consider data crawled before September 12, 2016. For some of the analysis performed in this paper we leveraged the Hatebase API¹¹. Hatebase is a repository of crowdsourced hate speech terms, which was useful to us when quantifying the amount of hate speech present on 4chan. **Ethical considerations.** Dealing with online data raises ethical concerns, especially when the topics of discussion are often sensitive like in the case of 4chan. Although 4chan users are anonymous by design, looking at the activity generated by links on 4chan to third party services deanonymize 4chan users, which goes against the reasonable expectation of privacy inherent on 4chan. To treat data ethically, we followed the guidelines by Rivers et al. [20]. In particular, we did not further deanonymize 4chan users based on the information obtained via the other datasets under our control. Since the main dataset we used is entirely anonymous due to 4chan’s design, and we make no effort to break that anonymity, we believe

⁷<https://www.theguardian.com/technology/2016/oct/04/4chan-website-financial-trouble-martin-shkreli>

⁸<http://knowyourmeme.com/memes/he-does-it-for-free>

⁹<https://github.com/4chan/4chan-API>

¹⁰Escaped HTML.

¹¹<https://www.hatebase.org/>

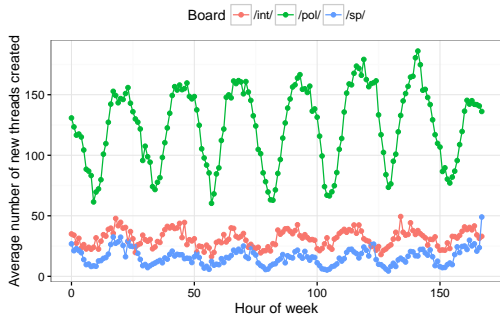


Figure 2: The average number of new threads created per hour of the week.



Figure 4: Heat map of the number of new /pol/ threads created per country, normalized by Internet-using population.

that the ethical concerns linked to this project are minimal.

Finally, we want to make an explicit mention about the content we are dealing with in this paper. /pol/ is, quite simply put, not a nice place, and the content posted by its users is distasteful at best, and often highly offensive. However, we believe that being open and honest with this work has legitimate scientific merit and we have chosen not to censor any content. With this said, we want to warn the reader that the remainder of this paper features images and language that is likely to be upsetting or uncomfortable.

3. GENERAL CHARACTERIZATION

In this section, we perform a general characterization of /pol/. In certain cases, we compare /pol/ to /sp/ and /int/, finding differences in the use of “permanent” identities and moderation behavior.

3.1 Posting Activity in /pol/

To begin understanding the behavior of /pol/ users, our first step is to perform a high-level examination of posting activity. To get an idea of how active 4chan users are on the different boards in our dataset, we first plot the average number of new threads created per hour of the week in Figure 2. The difference between the boards is immediately apparent: /pol/ users create an order of magnitude more threads than /int/ and /sp/ users at nearly all hours of the day.

Figure 4 shows a heat map of the number of new threads created per country, normalized by the country’s Internet-using population¹². Although the US dominates in terms of total thread creation (visible in the clear diurnal patterns from Figure 2), Australia, Canada, the UK, and the Scandinavian countries are all over-represented in terms of new threads per-capita. Even though 4chan

¹²Internet using population data from <http://www.internetlivestats.com/internet-users/>.

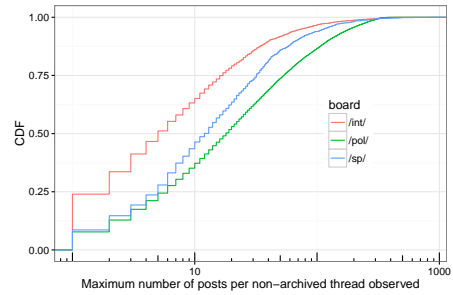


Figure 5: CDF of the maximum number of replies observed via the catalog for non-archived threads (likely removed by janitors).

is primarily an English speaking board, and indeed nearly every post in /pol/ is in English, we still find that decidedly non-English speaking countries are well represented: for instance, France, Germany, Spain, Portugal, and Eastern European countries like Serbia are all well represented. In other words, while /pol/ might be an ideological backwater, it is surprisingly diverse in terms of international participation.

Next, in Figure 3, we plot the distribution of the number of posts per thread on /pol/, specifically, reporting both the cumulative distribution function (CDF) and the complementary CDF (CCDF). We observe that the median is 7.0 and the mean of 38.4, i.e., the distribution is very skewed to the right, thus indicating that there are a few threads with a significantly higher number of posts. Note that this is as skewed as it potentially could be, due to 4chan’s bump limit system (see Section 2.1). Looking at both the CDF and the CCDF, the effects of the bump limit can be seen for threads with over 300 posts. The bump limit is designed to ensure that fresh content is always available, and based on the Figure 3 it appears to be doing its job: extremely popular threads are only able to get so popular before they are dropped from the catalog allowing fresh content to rise to the top.

Considering 4chan’s generally lax moderation, an important question arises: how much content violates the (few) rules of the board? To this end, in Figure 5, we plot the CDF of the maximum number of replies per thread observed via the /pol/ catalog, but for which we later receive a 404 error when trying to retrieve the archived version. While we believed these threads are most likely to have been deleted by a janitor, they could also be due to issues with 4chan’s servers. Another potential cause could be that a janitor moved a thread from one board to another, although anecdotally, we tend to see threads moved to /pol/. Somewhat surprisingly, there are many “popular” threads that are deleted, as the median number of posts in a deleted /pol/ thread is around 20, as opposed to 7 for the threads that are successfully archived. For /int/, the median number of posts in a deleted thread (5) is quite a bit lower than the median number of posts in archived threads (12). This difference is likely due to a combination of two things: 1) /int/ moves much slower than /pol/, giving moderators enough time to delete violating threads before they become overly popular, and 2) /pol/’s relatively lax moderation policy allows borderline threads to grow for awhile before they get out of control.

3.2 Tripcodes, Poster IDs, and Replies

We now focus on analyzing the use of tripcodes and poster IDs (introduced in Section 2.1) on 4chan, aiming to shed light on 4chan’s user base. Naturally, this a non-trivial task, since, due to the site’s anonymous and ephemeral nature, it is hard to build a unified network of user interactions. However, we can indeed leverage 4chan’s pseudo-identifying attributes (namely, tripcodes and poster

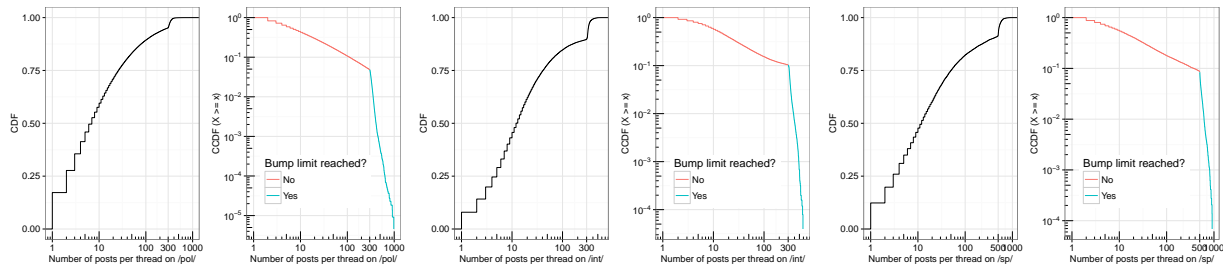


Figure 3: Distributions of the number of posts per thread on /pol/ (note log-scale on x-axis). We plot both the CDF and CCDF to show both the typical thread as well as threads that reach the bump limit. Note that the bump limit for /pol/ and /int/ is 300 at the time of this writing, while for /sp/ it is 500.

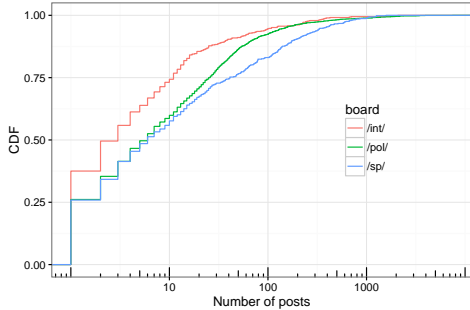


Figure 6: CDF of the number of posts per unique tripcode.

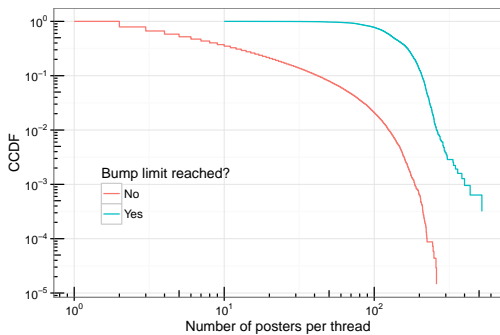


Figure 7: CCDF of the number of unique posters per thread on /pol/. We separate the distribution into threads that did not reach the bump limit vs. those that did reach it.

IDs) in order to get a first glimpse at both micro-level interactions and individual poster behavior over time.

Overall, we find 188,849 posts with a tripcode attached across /pol/ (128,839 posts), /sp/ (42,431), and /int/ (17,578). Note that unique tripcodes do not necessarily correspond to unique users, since users can use any number of tripcodes they desire. Figure 6 plots the CDF of posts per unique tripcode, for each of the three boards we study, showing that the mean and median are 36.08 and 6.50, respectively. We observe that 25% of tripcodes (over 30% on /int/) are only used once, and that, although /pol/ has many more posts overall, /sp/ has more active “tripcode users” – specifically, about 17% of tripcodes on /sp/ are associated to at least 100 posts, compared to about 7% on /pol/.

The closest we can get to knowing how unique users are engaged in 4chan threads is via poster IDs. Unfortunately, we only realized that poster IDs were not made available via the 4chan JSON API once a thread is archived after we began crawling. However,

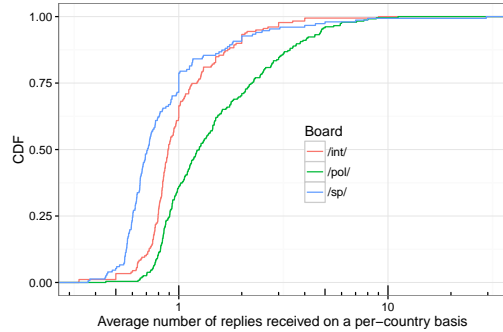


Figure 8: Distribution of the average number of replies received per country, per board.

the HTML version of archived threads *do* include poster IDs, and thus we additionally began collecting the HTML version of threads starting on August 17, 2016. In the end, we collected the HTML for the last 72,725 (about 33%) of threads in our dataset.

Figure 7 plots the CCDF of the number of unique users per /pol/ thread, broken up into threads that reached the bump limit and those that did not. The mean and median number of unique posters in threads that reached the bump limit was 139.6 and 134.0, respectively. For typical threads (those that did not reach the bump limit), the mean and median is 14.76 and 5.0 unique posters per thread. Clearly, even though 4chan is anonymous, the most popular threads have many voices.

4chan does not have the functionality to reply to a particular post in the traditional fashion. Instead, users can reference another post number N by putting $\gg N$ in their post text. The standard 4chan.org UIs then treat it as a reply (see Figure 1 for examples). Note that this is a different thing than just posting in a thread (which might also be considered a “reply” to the OP). Here, users are *directly* replying to a specific post (not necessarily the OP) and must go out of their way to do so. There are some caveats, for example, you can reply to the same post multiple times and you can also reply to multiple posts at the same time. With that caveat in mind, we can exploit the reply functionality to get an idea on how engaged users are with each other.

First, we find that about 57% of posts never receive a direct reply across all three boards (57% in /pol/, 49% in /int/, and 60% in /sp/). Taking the posts with no replies into account, we see that /pol/ ($\mu = 0.83$) and /int/ ($\mu = 0.80$) have many more replies per post than /sp/ ($\mu = 0.64$), however, the standard deviation on /pol/ is much higher ($\sigma = 2.55, 1.29, \text{ and } 1.25$ for /pol/, /int/, and /sp/, respectively).

Next, Figure 8 plots the CDF of the average number of replies re-

/pol/		/int/		/sp/	
Country	μ Replies	Country	μ Replies	Country	μ Replies
China	1.57	Thailand	1.13	Slovenia	0.91
Pakistan	1.42	Algeria	1.12	Japan	0.84
Japan	1.35	Jordan	1.04	Bulgaria	0.81
Egypt	1.33	S. Korea	1.02	Sweden	0.75
Tri. & Tob.	1.28	Ukraine	1.00	Israel	0.74
Israel	1.27	Viet Nam	0.97	Argentina	0.72
S. Korea	1.20	Tunisia	0.97	India	0.72
Turkey	1.18	Israel	0.97	Greece	0.72
UAE	1.20	Hong Kong	0.92	Puerto Rico	0.70
Bangladesh	1.15	Macedonia	0.91	Australia	0.68

Table 2: The top 10 countries (with at least 1,000 posts) in terms of direct replies received per post for each board in our dataset.

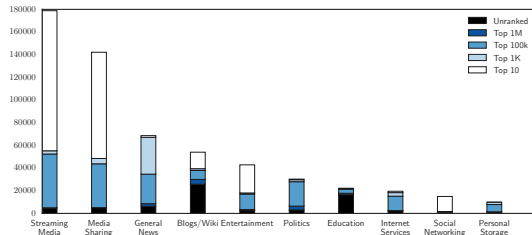


Figure 9: Distribution of different categories of URLs posted in /pol/, together with the alexa ranking of their domain. As it can be seen, some categories tend to be heavily composed of very popular domains, while others have a strong representation of domains that are not among the top popular ones.

ceived per poster per board, aggregated by the country of the poster. I.e., it is the distribution of mean replies received per country. From the figure we see that there are pretty substantial differences between the different boards. Thus while /pol/ posts are, *on average*, likely to receive more replies than /sp/ and /int/, the distribution is heavily skewed.

In Table 2 we show the countries (with at least 1,000 posts) in the top 10 of average replies received per post for each of the boards in our dataset, which lets us zoom in a bit on the tails in Figure 8. Every single country in the top 10 from /pol/ has more average replies than every other country in /sp/ and /int/. Overall there is minimal overlap between the boards: while Israel is in the top 10 of all three, only two other countries (Japan and South Korea) appear in the top 10 list of more than one board.

4. CONTENT ANALYSIS

In this section, we present an exploratory analysis of content posted on 4chan. First, we look at media shared on /pol/, then, we cluster content based on its geo-political nature.

4.1 Media

4.1.1 Links

Unsurprisingly, users on /pol/ often post links to external content, e.g., to share and comment on news and event. As we show in Section 5, users also do so in order to identify and coordinate possible targets for attacks and hatred on other platforms.

We study the nature of the links (URLs) posted on /pol/, relying on the McAfee SiteAdvisor service¹³. Provided with a URL, this service returns its category (e.g., “Entertainment” or “Social Networking”). We also aim to understand to what extent /pol/ users

¹³<https://www.siteadvisor.com/>

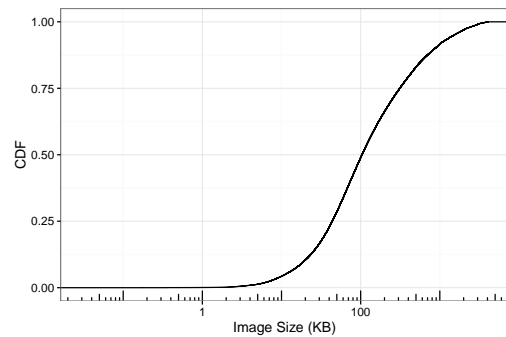


Figure 10: Distribution of file sizes for unique images posted to /pol/.

post links to popular websites, compared to less known ones – based on the domain’s Alexa ranking.¹⁴

In Figure 9, we plot the distribution of categories of URLs posted in /pol/: “Streaming Media” and “Media Sharing” are the most common ones on the board, with YouTube playing a key role. Overall, we observe that URL categories that are less common on /pol/ tend to include links to less popular websites as per Alexa ranking.

The single most popular site on /pol/ is YouTube, with over an order of magnitude more URLs posted than the next two sites, Wikipedia and Twitter. Next is Archive.is, a site that lets users take on demand “snapshots” of a website. On /pol/, it is often used to record content such as tweets, blog posts, or even news stories that users think might get deleted soon after sharing them on 4chan. Then Wikileaks and pastebin, both info-dump sites. Next, DonaldJTrump.com appears, followed by dailymail and Breitbart, both news outlets of questionable reputation. Rounding out the top 10 most popular sites on /pol/ is archive.org, another web page “snapshot” system. We find it somewhat telling that “legitimate” news sites like the Telegraph, BBC, and Guardian do appear outside the top 10 most common domains on a forum supposedly focused around politics and current events.

4.1.2 Images

While 4chan generates large amounts of original content, there clearly is content that is reposted – in fact, memes are, almost by definition, going to be posted numerous times. To this end, we focus on images, aiming to measure to what extent they are reposted from other boards.

In our dataset, we observe 1,003,785 unique images out of a total 2,210,972 images posted on /pol/, corresponding to almost 800GB. Figure 10 plots the CDF of sizes for unique images uploaded to /pol/. The median and mean size of unique images uploaded is 103.9 KB and 321.3 KB, respectively.

Using the image hash as a unique identifier, Figure 11 plots the CCDF of the number of posts each unique image appears in. While we note that the figure should be considered a *lower* bound on image reuse (it only captures *exact* reposts), we can see that the majority (about 70%) of the 1,003,785 images posted in our dataset are posted only once; in fact nearly 95% of images are posted no more than 5 times. That said, there is a very long tail. The most popular image (Figure 12) was posted 838 times and depicts a (skeptical?) Pepe, a meme recently declared a hate symbol by the Anti-Defamation League [13]. While Figure 12 is clearly the most common of Pepees, we have included a collection of somewhat rarer Pepees in Section 9.

Next, we investigate how many how many images have been pre-

¹⁴<http://www.alexa.com/>

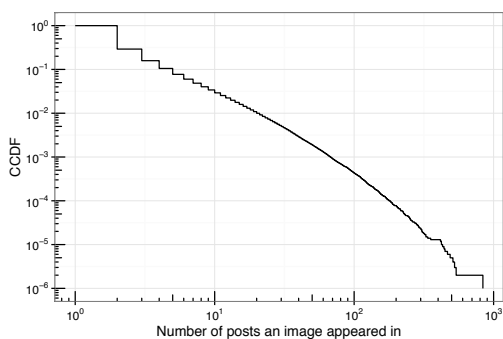


Figure 11: CCDF of the number of posts exact duplicate images appeared in on /pol/.

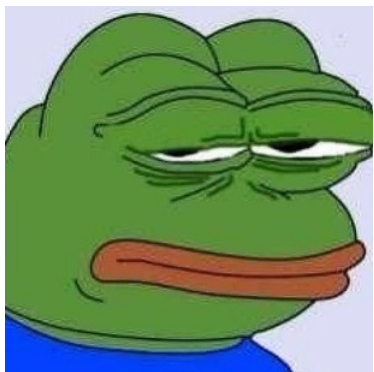


Figure 12: The most popular image on /pol/ during our collection period. Perhaps the least rare Pepe.

viously posted on any 4chan board. When posting an image on 4chan, although the filename the image was uploaded with still appears along with the post, the file when served is renamed to the 13-digit Unix epoch (in milliseconds) corresponding to when the post was made. Therefore, we assume that, if the filename the image was uploaded with is a 13-digit string, and this corresponds to an epoch occurring after 4chan was founded, then the user is likely to have downloaded it from 4chan.¹⁵ Using this heuristic, we find that about 31% of images (683,713 of 2,210,972 total) on /pol/ have been previously posted on 4chan. For /sp/ and /int/ about 26% (56,167 of 219,625) and 30% (100,686 of 330,582) of images were previously posted on 4chan, respectively. Figure 13 plots the distribution of the difference in time between when an image is re-posted to each board in our dataset and the time it is originally posted to 4chan. We observe a 41.3 days median “delay” for reposted images on /pol/, with about one quarter of re-posted images appearing on 4chan within the previous week. For /sp/ and /int/, the medians are 118.7 days and 84.1 days, respectively.

From these numbers we can draw the following conclusions. First, as a whole, each of the boards in our dataset produce a surprising amount of “original” content. Even with an extremely conservative estimation, /pol/ users posted over 1M unique images, the majority of which were either completely original content or sourced from a platform other than 4chan. Next, we see that across all three boards, that there is a substantial amount of re-use *within* 4chan. This makes sense considering that 4chan is relatively famous for memes, and a meme is only a meme if it is seen many

¹⁵This is certainly just a heuristic and there are many places besides 4chan images could have been acquired, however, even a casual browsing of 4chan will make it clear how common this behavior is.

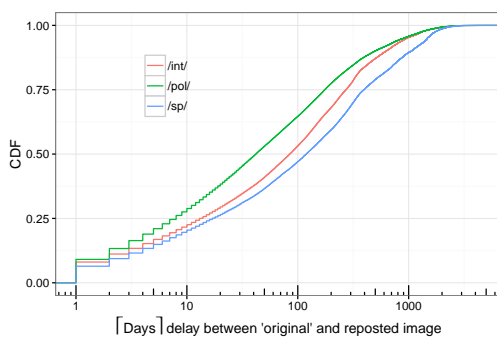


Figure 13: CDF of the delay between an image originally appearing on 4chan and it being reposted on /pol/, /sp/, /int/.

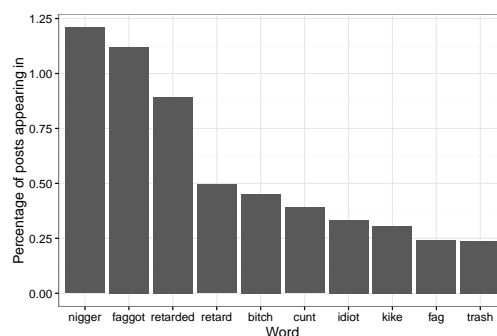


Figure 14: Percentage of posts on /pol/ that the top 10 most popular hate words appear in.

times. Overall, these results point to the fact that the constant production of new content is likely one of the reasons that /pol/ is at the heart growing hate movement on the Internet.

That said, a major limitation of this facet of our study is that we do not compare image contents. As future work, we believe that there is value in mapping the progression of memes. For example, there are numerous Pepe images, that clearly are all within the same “family” of memes, however, classifying them as such is a machine learning problem beyond the scope of this work.

4.2 Hate Around the World

/pol/ is a pretty hateful place, however, quantifying hate is an open problem. To get a first idea of how prevalent hate speech is on /pol/, Figure 14 plots the percentage of /pol/ posts that the top 10 most “popular” hate words from the hatebase dictionary appear in. “Nigger” is, by far, the most “popular” hate word used with “faggot” as a close second. Rounding out the top 10 is “trash.” In general usage, this is obviously not a hate word, however, /pol/ tends to use it in a disparaging fashion (e.g., calling an ethnic group “trash”).

Figure 15 plots a heat map of the percentage of posts that contain hate speech per country with at least 1,000 posts on /pol/. Countries are placed into seven equally populated bins and colored from blue to red depending on the percentage of their posts contain a hate word from the hatebase dictionary.

First, we note that there are clear differences in the use of hate speech by different countries: ranging from around 3% (e.g., Vatican City) to around 20% of posts (e.g., Cyprus). The majority of countries exhibit hate speech in between 8% and 12% of their posts, however. We also see elevated uses of hate speech in certain European countries (e.g., Italy, Spain, Greece, and France) that possibly

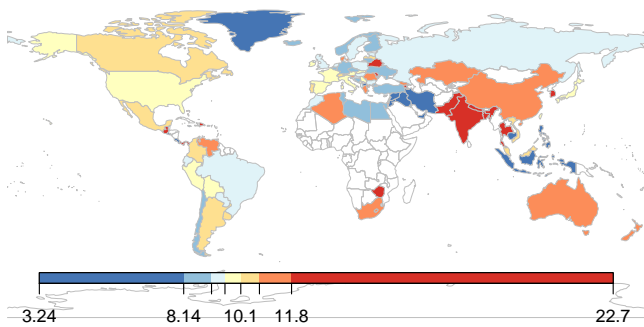


Figure 15: Heat map showing the percentage of posts with hate speech per country.

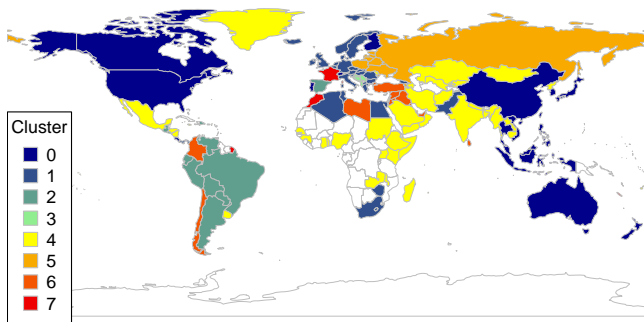


Figure 16: Worldmap colored by content analysis based clustering. [Best viewed in color.]

has to do with the current immigration crisis. Finally, we note that the Anglosphere countries are generally close to each other with respect to hate speech usage: on the low end we see just over 9% in the UK and on the high end, a bit over 11% for Australia and South Africa.

Now, to get an idea of how /pol/’s conversations are related to the country of the posters, we perform some basic text classification aiming to evaluate whether or not different parts of the world are talking about “similar” topics. We first group posts together based on their country as returned by the 4chan API. After removing stop words and performing stemming, we build term frequency-inverse document frequency (TF-IDF) vectors for each country. These vectors represent the frequencies with which different words are used by each country, but are down-weighted by the general frequency of each word across all posts. Finally, we apply spectral clustering over the vectors using the Eigengap heuristic to automatically identify the number of target clusters. In Figure 16, we present a world map colored according to the 7 clusters generated.

The plot makes it pretty clear that posters from the same region of the world tend to say similar things. However, there are a few interesting things to point out. First, the clustering clearly identifies a core sector of the Balkans near Serbia. Next, we note that while the majority of Western Europe falls into the same cluster, France and Spain are notably in different clusters. France is in a cluster with Morocco and former French colonies such as New Caledonia and Martinique. Spain is in the cluster including the majority of South American countries. Oddly enough, while Spain is grouped together with its former colonies for the most part, it is also clustered with Brazil, while Portugal is clustered with the United States, Canada, and Australia. In addition to Portugal’s odd clustering, we

also believe its worthwhile to note that Finland is in this same cluster.

5. RAIDS TOWARDS OTHER ONLINE SERVICES

As we saw /pol/ is not only used as a self-contained discussion board, but its users frequently post links to the rest of the web (Section 4.1.1). Some of these links are just posted as a commentary to the discussion, but some others serve to call for /pol/’s users to act upon, in what we call “raids.” A raid is, broadly speaking, an attempt to disrupt some other site. A raid differs from a DDoS in that the goal is not to disrupt the service from a network perspective (i.e., to make it unavailable), but rather to disrupt the *community* that calls the service home.

Raids on /pol/ are semi-organized. The prototypical raid would be a thread or post made with a link to a target and perhaps the text “you know what to do.” Other 4chan users would then harass the target, for example on Twitter or YouTube. The 4chan thread itself can then become an aggregation point where users do things like post screen shots of the target’s reaction, share the sock puppet accounts they are using to harass, and discuss particular things to say to the target. Unlike 4chan’s earliest days, raids are now strictly prohibited throughout the site, and special mention is made on /pol/’s rules as well, however, we believe there is evidence that they still occur.

In this section we aim to understand the way that raids work on 4chan. We thus first begin with a case study of a very broad-target raid. Next, we find large scale evidence of raids as well as providing a basic algorithm for detecting a raid taking place.

5.1 Case Study: “Operation Google”

Early morning September 22, 2016 Figure 17 was posted to /pol/. In it, a poster calls for the execution of “Operation Google,” a response to Google’s announced anti-trolling machine learning technology¹⁶. /pol/ users theorize that they can poison Google’s anti-trolling technology by using, e.g., “Google” instead of “nigger” and “Skype” instead of “kike.” In particular, the Operation Google post calls for using these phrases to disrupt social media sites like Twitter. By examining the impact of Operation Google on both /pol/ and Twitter we can gain useful insight into just how “efficient” /pol/ is in spreading their “ideology.”

Figure 18 plots the normalized usage of the specific replacements called for in the Operation Google post. The effects within /pol/ are quite evident: on September 22nd we see the word “Google” appearing at over 5 times its normal rate, while “Skype” appears at almost double its normal rate. To some extent, this illustrates how quickly /pol/ can execute on a raid, but also how short of an attention span its users have: by September 26th the burst in usage of Google and Skype had died down. While we still see elevated usages of Google and Skype, it is quite worthwhile to note that there is no discernible change in the usage of “nigger” or “kike.” It remains to be seen how long this elevated usage will persist. Clearly, it has not taken over completely, but it seems to have become a part of /pol/’s vernacular.

While it is quite clear that /pol/ seized upon the concept of Operation Google, a larger question remains: what were the effects *outside* of /pol/. To that end, we counted tweets in our dataset of over 60M tweets that contained any of the hashtags listed in Figure 17, namely #worthlessgoogs, #googlehangout, #googleriots, #googlesgonnagoog, and #dumbgoogles. Figure 20 shows examples of such tweets.

¹⁶<https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/>

Operation Google

"the new nigger"

If you've been seeing this around /pol/ you're probably wondering, "why the fuck would we say Google instead of nigger? It's dumb and cringey as fuck", well to answer this it's done after Google's recent announcement to censor certain words looked through their search engine, just to keep safe spaces extra "safe". Our response is to make it so Google would have to censor their own company by making them a racial slur towards blacks.



"That'll never work, its too hard!"

Don't doubt the power of Kek and meme magic. Look at Pepe, Triggypuff, or Carl the Cuck, the latter being chosen simply based off a dubs post. If we simply call a bunch of black people "Googles" they will 100% be anally devastated. Remember these are the same people that riot over a nameless gang-banger being shot, their skin is fragile and easily penetrated.

"Too reddit for my tastes"

This is hardly close to reddit and if you unironically think that, then maybe you need to go back to reddit.

"Well okay, then how do we get this around?"

Social media, social media, social media. Get this fucking shit trending to the fucking moon. Recently there has been riots in Charlotte so BLM tweets will be nice and fresh for next couple of days. What you do then is respond to these people calling them "stupid fucking googles" or "worthless google, kill yourself" try to make it sound as venomous as possible so they definitely get triggered. Also include some hashtags I'd recommend:

- #WorthlessGoogs
- #GoogleHangout (include pics of googs being hanged)
- #GoogleRiots (whenever some dumb BLM riot happens)
- #GooglesGonnaGoog
- or #DumbGoogles

Also call Jews-Skypes

GOD SPEED ANONS



Figure 17: An image describing /pol/'s "Operation Google."

Figure 19 shows that #dumbgoogles and #googleriots first appeared on the 22nd of September (after Operation Google was launched on 4chan). Other hashtags showed up later. This indicates that an attempt was indeed made to instigate censorship evasion on Twitter, as discussed earlier. However, a look at the percentage of tweets containing those hashtags shows that the impact of Operation Google was no where near as effective on Twitter as it was within /pol/ itself. For instance, at its highest peak (Figure 19), #dumbgoogles was found in only 0.00015% of the 23rd September tweets in our dataset, that is, only five tweets contained the hashtag #dumbgoogles out of a total of 3,343,941 tweets on that day.

Thus, while we definitely do see evidence of Operation Google in effect on Twitter, it seems to be primarily contained within /pol/ at this time. This seems at odds with the level of media coverage Operation Google has received.¹⁷ Although it would seem that Operation Google was a failure, we believe that based on media coverage alone it served as successful propaganda for /pol/.

5.2 YouTube Comments

While Operation Google is certainly an interesting case study, raids are a larger problem in general. While services and researchers are certainly starting to take online harassment and trolling seriously, very little work has looked at how the trolls oper-

¹⁷E.g., <http://www.telegraph.co.uk/technology/2016/10/03/internet-trolls-replace-racist-slurs-with-online-codewords-to-av/>.

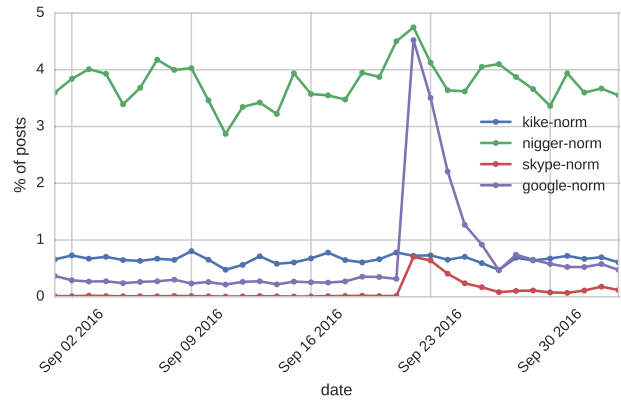


Figure 18: The effects of Operation Google within /pol/.

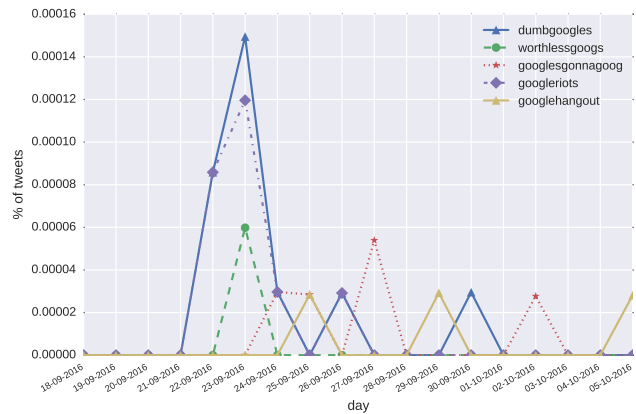


Figure 19: The effects of "Operation Google" on Twitter.

ate. We believe that understanding how forces *completely outside the control* of target services organize and execute their campaigns is crucial to mitigating what is quickly turning out to be a social menace.

YouTube is both the most popular site linked on /pol/ and also experiencing a big enough issue with their comments that they recently announced the controversial YouTube Heroes program.¹⁸ Thus, we examine the comments from 19,568 YouTube videos linked to by 10,809 /pol/ threads to look for raiding behavior at scale.

Unfortunately, finding evidence of raids is not an entirely straight forward task: explicit calls for raids have been an offense that can lead to the user being banned on /pol/ for some time. Instead of looking for a particular trigger on /pol/, we instead look for elevated activity in YouTube comments that /pol/ links to. What we believe happens is that a /pol/ user will see a YouTube video linked, and be incited into attacking either the subject of the video or perhaps other YouTube commenters. In typical raid fashion, they might even report back to the /pol/ thread in some fashion.

One way this behavior might manifest itself is with bursts of activity. For example, if the comments to a YouTube video experience a peak in commenting activity within the lifetime of the 4chan thread it was linked to, it is an indication that a raid might be occurring.

Let us consider a /pol/ thread x and the comments to YouTube

¹⁸<https://youtube.googleblog.com/2016/09/growing-our-trusted-flagger-program.html>



Figure 20: Two tweets featuring Operation Google hashtags in combination with other racist memes.

video y that was linked to in x . Let us further say that $x(t)$ is the set of timestamps of posts in x , and $y(t)$ is the set of timestamps of the comments associated with y . Because the lifetime of threads on /pol/ is quite dynamic, we shift and normalize the time axis for both $x(t)$ and $y(t)$ such that $t = 0$ corresponds to the moment when y was mentioned and $t = 1$ corresponds to the last post in the /pol/ thread:

$$t \leftarrow \frac{t - t_{yt}}{t_{last} - t_{yt}}. \quad (1)$$

In other words, we normalize to duration of the /pol/ thread’s lifetime, with the qualification that we consider only /pol/ posts that occur after the YouTube mention, and consider only YouTube posts that occurred within the (normalized) $[-10, +10]$ period (35% of total DB).

We can use a relatively simple method for detecting peaks in YouTube commenting activity relative to /pol/ threads:

1. We extract the timestamp of each link to YouTube that appears on /pol/, as well as the timestamp of the /pol/ post that it appeared in (the *mentioning* post).
2. For each corresponding YouTube comment thread, we calculate the Probability Density Function (PDF) of comment arrivals by means of Kernel Density Function estimator.
3. We say a peak has occurred when the slope of the PDF

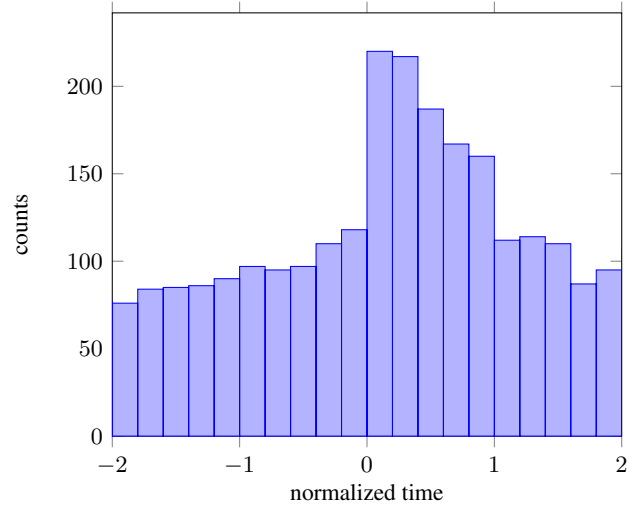


Figure 21: Distribution of the distance (in normalized thread lifetime) of the highest peak of activity in YouTube comments and the /pol/ post they appeared in. ($\frac{1}{5}$ unit bins)

changes from positive to negative and we take the highest one.

4. Next, we count number of YouTube comment threads for which the highest peak to the corresponding mention on 4chan occurs, per 1 unit (normalized) time bin.

Figure 21 plots the distribution of the distance between the highest peak in YouTube commenting activity and the /pol/ post the YouTube video was mentioned in. We see that 14% of the YouTube videos experienced a peak in activity during the time they were mentioned on /pol/ and the time the thread they were mentioned in died. In other words, it seems that either /pol/ users are in fact commenting on the YouTube videos linked (i.e., raiding them), or, alternatively, that they are linking YouTube videos that are currently undergoing a peak in commenting activity completely independent of /pol/.

If raiding behavior is taking place, then the comments on both /pol/ and YouTube are likely “synchronized” somehow (in a broad sense). In a “perfect” world, where /pol/ posters had perfect multi-tasking and completely took over a YouTube video’s comments, we might expect the synchronization to be perfect. I.e., there would be zero lag between the time series formed from the /pol/ posts and the YouTube comments /pol/ was attacking.

We can make use of cross-correlation to compute the synchronization lag. Cross-correlation is a simple but powerful tool mainly used to estimate the lag between two signals¹⁹ or search for a specific pattern within a signal. The method is also known as the sliding inner-product, since the basic idea is to slide one signal with respect to the other and calculate the dot product between the two signals for each possible lag. The estimated lag between the two signals is the one that maximizes the dot-product (also called maximum matching).

There are some requirements we must meet to apply this method to sequences of events like /pol/ and YouTube comments. First of all, we have to represent the sequences as signals. The most common and powerful way to represent sparse events is by means of Delta Dirac functions. This function gives us the tools to formalize the problem in very simple and elegant way.

¹⁹Under some weak conditions, the cross-correlation method have been proved to be optimal in lag estimation.

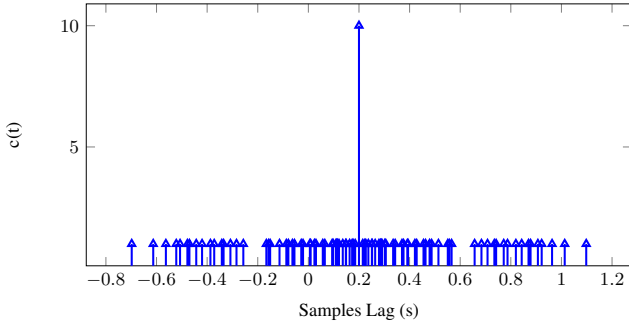


Figure 22: Cross-correlation between shifted sequences where $y(t) = x(t - 0.2)$.

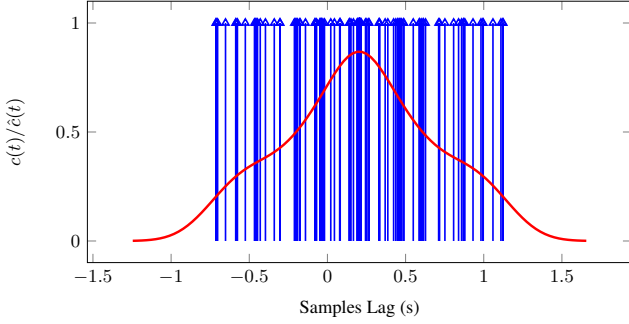


Figure 23: Cross-correlation between sequences whose samples have slightly different lags

Thus, let us expand $x(t)$ and $y(t)$ into trains of Dirac delta functions::

$$\begin{aligned}
 x(t) &= \sum_{i=1}^{N_x} \delta(t - t_x^i) \\
 y(t) &= \sum_{j=1}^{N_y} \delta(t - t_y^j)
 \end{aligned}
 \tag{2}$$

This mathematical representation lets us calculate $c(t)$, the continuous time cross-correlation between the two series²⁰:

$$\begin{aligned}
 c(t) &= y(t) \otimes x(t) = y(t) * x(-t) = \\
 &= \int_{-\infty}^{\infty} x(t + \tau) y(\tau) d\tau = \\
 &= \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \delta\left(t - (t_y^j - t_x^i)\right)
 \end{aligned}
 \tag{3}$$

where \otimes and $*$ denote respectively the cross-correlation and convolution operators. The resulting cross-correlation function is a delta Dirac train as well, representing the set of all possible inter-arrival times between elements from the two sets.

In order to give the reader a better intuition about the method, we will show two simple examples. If we consider the simple case in which $y(t)$ corresponds to a shifted version of $x(t)$, that is $y(t) = x(t - \Delta T)$, it is easy to see that, as in the continues time case, $\arg \max_t(c(t)) = \Delta T$ (the high pulse in Figure 22 is due to

²⁰Note that this is different from the discrete-time cross-correlation, since the samples are not equally spaced.

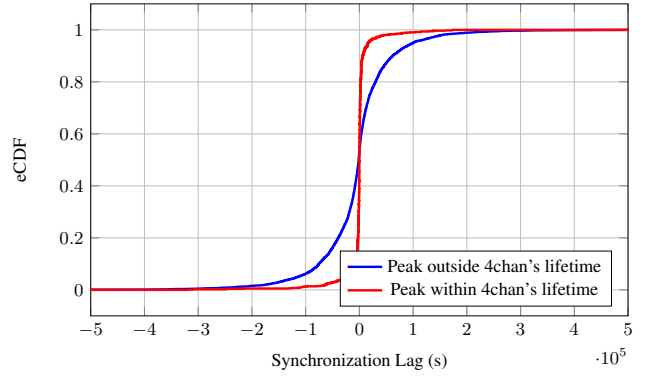


Figure 24: CDF of estimated synchronization lag between /pol/ threads and corresponding YouTube comments. We compare all threads to just those threads with a corresponding peak in YouTube activity during the thread's lifetime.

the perfect overlapping of the two pulse's trains). Now let us consider the case where $y(t)$ is again a shifted version of $x(t)$, but this time each sample is delayed with a slightly different time lag, that is: $y(t) = \sum_{i=1}^{N_x} \delta(t - (t_x^i + \Delta T + \Delta t^{i,j}))$, where $\{\Delta t^{i,j}\}$ are samples of some zero-mean process. In this case we cannot find a lag that gives us a perfect match of the two sequences (e.g. the pulses in Figure 23 have all the height), however the $c(t)$ function will be characterized by a high concentration of pulses around ΔT (Figure 23). As we did for peak activity detection, we can estimate the more likely lag by computing the associated PDF function $\hat{c}(t)$ by means of the Kernel Density Estimator method, and then compute the global maximum:

$$\begin{aligned}
 \hat{c}(t) &= c(t) * k(t) = \\
 \hat{\Delta T} &= \arg \max_t \hat{c}(t)
 \end{aligned}
 \tag{4}$$

where $k(t)$ is the kernel smoothing function (typically a zero-mean gaussian function).

It is worth pointing out that, from a mathematical point of view, this is equivalent to computing the cross-correlation between the PDF functions associated to $x(t)$ and $y(t)$ series.

$$\begin{aligned}
 \hat{c}(t) &= c(t) * k(t) = \\
 &= (y(t) * x(-t)) * (k_x(t) * k_y(t)) = \\
 &= (y(t) * k_y(t)) * (x(-t) * k_x(t)) = \\
 &= \hat{y}(t) * \hat{x}(-t)
 \end{aligned}
 \tag{5}$$

where we have exploited the fact that kernel function $k(t)$ can always be expanded in the convolution of two other kernels $k_x(t)$ and $k_t(t)$ (e.g. the convolution of two gaussian functions is a gaussian funtion).

Now, Figure 24 plots the CDF of estimated lags as computed via cross-correlation between /pol/ threads and corresponding YouTube comments. We specifically compare the distribution for *all* /pol/ threads/YouTube comments vs. just the /pol/ thread/YouTube comments pairs where a peak in activity on YouTube was detected during the /pol/ thread's lifetime (i.e., there was a peak on $[0, 1]$ in Figure 21). From Figure 24 we can see that when a peak in YouTube comment activity was detected during the /pol/ thread that mentioned it we tend to see drastically lower lag.

To provide some intuition behind the implications of this, Figure 25 plots the distribution of the distance between the high-

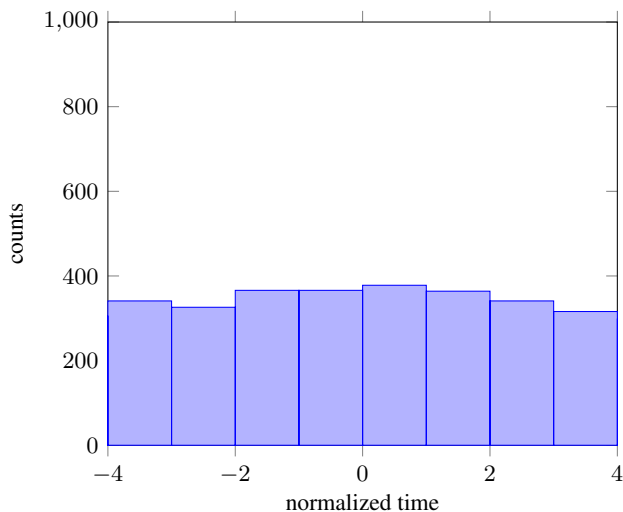


Figure 25: Distribution of the distance (in normalized thread lifetime) of the highest peak of activity in YouTube comments and the /pol/ post they appeared in. Most synchronized threads removed.

est peak in YouTube commenting activity and the /pol/ post the YouTube video was mentioned in, but with out any of the threads that appear in the middle 60% of the candidate threads (those where the YouTube comments experienced a peak within the /pol/ thread’s lifetime) in Figure 24. I.e., it is Figure 21 without the most synchronized threads. Now, the distribution is flat. Thus, if we focus on the middle 60% of candidate threads, we have essentially removed all the background noise that comes from incidental overlap in activity on both services.

We have thus far only analyzed activity from the point of temporally related events, and while it seems very likely that there are /pol/ users posting on YouTube videos they discover on /pol/, this is not sufficient evidence for raids. The /pol/ posters could just be commenting as might be expected of upright netizens. However, we can validate that raiding is taking place by looking at the contents of the YouTube comments.

Figure 26 plots the relationship between the number of hateful comments on YouTube (as determined by containing at least one word from the hatebase dictionary) and the synchronization lag between the /pol/ thread and the YouTube comments. The trend is quite clear: as the synchronization lag between /pol/ and YouTube comments decreases the rate of hateful comments on YouTube increases.

As further validation, Figure 27 plots the CDF of estimated synchronization lag between /pol/ threads and YouTube comments. We separate YouTube comments that had more hate words *during* the /pol/ thread from those that had more hates *prior* to the /pol/ thread. I.e., we are comparing threads where we /pol/ appears to have had a negative impact with those where they did not. From the Figure we see that the YouTube comments with more hate speech during the /pol/ thread’s lifetime are significantly more synchronized with the /pol/ thread itself.

6. RELATED WORK

Although 4chan has received a large degree of interest in popular media [4, 5, 11, 24], to the best of our knowledge, there is very little work from the research community.

Bernstein et al. [6] study /b/, the “random” board on 4chan. /b/ was the first board on the site and remains the most active one – in

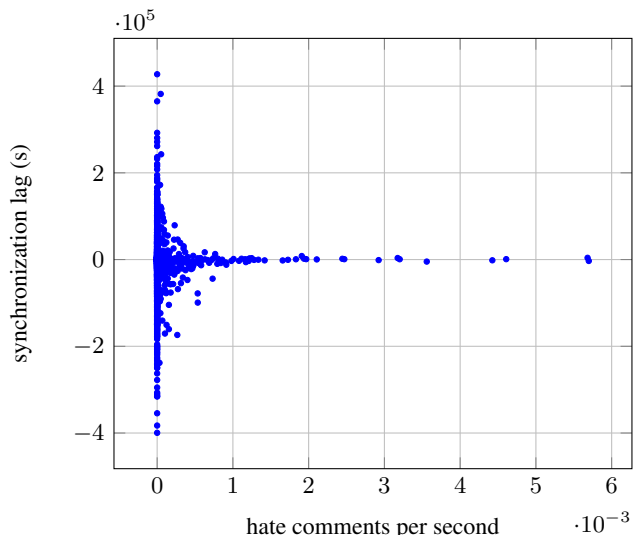


Figure 26: The relationship between hateful YouTube comments and synchronization lag between /pol/ threads and corresponding YouTube comments. Each point is a /pol/ thread. The x-axis is the number of hateful comments per second that occur within the /pol/ thread lifetime and the y-axis is the synchronization lag.

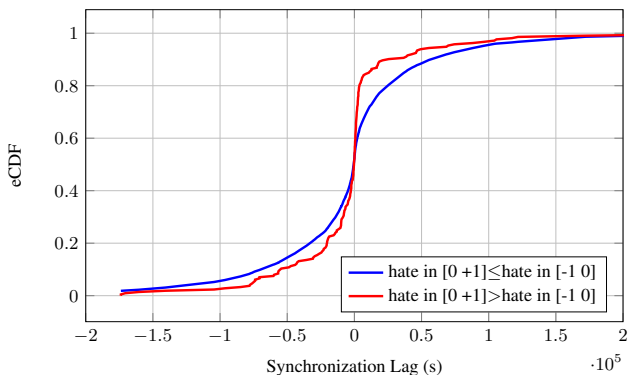


Figure 27: CDF of estimated synchronization lag between /pol/ threads and corresponding YouTube comments. We compare threads whose corresponding YouTube video contains more hate comments in the [0 +1] period (thread lifetime) rather than [-1 0]

fact, moot considers it to be the “life force” of 4chan [22]. Using a dataset of 5.5M posts from almost 500K threads collected over a two-week period, [6] primarily studies how /b/ works from the anonymity point of view. Similar to our work, they are also interested in understanding how 4chan’s “bump” system affects thread development and lifetime, finding that the median lifetime of a /b/ thread to be only 3.9 minutes (and 9.1 minutes on average).

In [17], Potapova et al. study the influence of anonymity on the level of aggression and usage of obscene lexis. They analyze the lexicon of different anonymous forums and social networks, mainly focusing on Russian-language platforms. They also include 4chan as the only English-speaking case study, and find no correlation between anonymity and aggression. They also point out that obscene language does not necessarily carry aggression. According to their results, 8.2% of messages contain obscene language although expressing positive emotions. In follow-up work [9, 18], authors use 4chan messages to evaluate automatic verbal aggression detection techniques.

Other work has looked at the analysis of social networks, other than 4chan, that are characterized by (semi-)anonymity and/or ephemerality. For instance, Correa et al. [8] study the differences between content posted on anonymous and non-anonymous social media, showing that linguistic differences between Whisper posts (anonymous) and Twitter (non-anonymous) are significantly different, and they train automated classifiers to discriminate them with reasonable accuracy (73%). Also, Onwuzurike and De Cristofaro [15] look at security and privacy issues in Android apps offering anonymity and ephemerality. Peddinti et al. [16] analyze users' anonymity choices during their activity on Quora, identifying categories of questions for which users are more likely to seek anonymity. They also perform an analysis of Twitter to study the prevalence and behavior of so-called "anonymous" and "identifiable" users, as classified by Amazon Mechanical Turk workers, and find a correlation between content sensitivity and a user's choice to be anonymous. Roesner et al. [21] analyze why people use Snapchat, performing a survey of 127 adults, and find that privacy is not the major driver of adoption, but the "fun" of self-destructing messages.

Finally, Hosseinmardi et al. [10] analyze user behavior on ask.fm (a social network where users create profiles and can send each other questions) by building an "interaction graph" between 30K profiles [10]. They characterize users in terms of positive/negative behavior, positive/negative in-degree/out-degree and analyze the relationships between these factors.

7. CONCLUSION

This paper presented the first large-scale study of /pol/, 4chan's politically incorrect board. We provided a general characterization, which compared activity on /pol/ to two other boards on 4chan, /sp/ ("sports") and /int/ ("international"). We first showed that each of the boards has different behavior with respect to thread creation and posts. In particular, we looked into the impact of 4chan's "bump limit" system on discourse, finding that it does indeed seem to result in "fresh" content on a consistent basis. We made use of the "flag" feature present on the three boards we examined to get an idea on the demographics of users, finding that while in absolute numbers Americans dominate the conversation, many other countries (both native English speaking and not) are well represented in terms of per-capita. We also showed differences in the "maturity" of threads with respect to moderator actions across the boards in our dataset.

We next examined the content posted to /pol/. We first found that, by far, the majority of links posted to 4chan are to YouTube. We also saw that /pol/ posts many more links to tabloid or questionable news outlets than main stream media sites. This is somewhat expected considering /pol/'s alt-right leanings. By looking at meta data associated with posted images we learned that most content on 4chan is reasonably unique: 70% of the 1M unique images in our dataset were posted only once and 95% were posted less than 5 times. /pol/'s ability to find and/or produce original content is likely one of the reasons it is thought to be at the center of hate on the Internet.

Finally, we studied raiding behavior via a quantitative case study of "Operation Google" as well looking for evidence of /pol/'s impact on YouTube comments. Although the impact of Operation Google on /pol/ itself was substantial, we found little evidence that it has spread to Twitter. However, we did find ample evidence of raiding activity on YouTube.

When looking for peaks of commenting activity on YouTube we discovered that they tend to occur within the lifetime of the thread they were posted to on /pol/. Because YouTube and 4chan are

completely separate platforms, we would expect to see that some of these activity peaks are background noise. However, we were able to filter out this background noise by using cross-correlation to estimate the synchronization lag between the /pol/ thread and the YouTube comments. Finally, we showed that as this synchronization lag approaches zero, there is a statistically significant increase in the number of hate words on YouTube. I.e., we provide quantitative evidence for raiding behavior.

Acknowledgments. Authors wish to thank Andri Ioannou and Despoina Chatzakou for helpful comments and feedback. This research is supported by the European Union's H2020-MSCA-RISE grant "ENCASE" (Grant Agreement No. 691025) and by the EP-SRC under grant EP/N008448/1. Jeremiah Onaolapo was supported by the Petroleum Technology Development Fund (PTDF), Nigeria.

8. REFERENCES

- [1] ALFONSO, F. I. After 4chan manhunt, cat-kicker slapped with animal cruelty charges. <http://www.dailydot.com/news/walter-easley-cat-kicker-animal-cruelty/>, 2016.
- [2] ANDERSON, N. 4chan tries to change life OUTSIDE the basement via DDoS attacks. <http://arstechnica.com/tech-policy/2010/09/4chan-tries-to-change-life-outside-the-basement-via-ddos-attacks/>, 2016.
- [3] BARNES, J. E. Europol Warns of Cybercrime Surge. <http://on.wsj.com/2cAWwhB>, 2016.
- [4] BARTLETT, J. 4chan: the role of anonymity in the meme-generating cesspool of the web. <http://www.wired.co.uk/article/4chan-happy-birthday>, 2016.
- [5] BBC NEWS. Stolen celebrity images prompt policy change at 4Chan. <http://bbc.in/2cKde3a>, 2016.
- [6] BERNSTEIN, M. S., MONROY-HERNÁNDEZ, A., HARRY, D., ANDRÉ, P., PANOVICH, K., AND VARGAS, G. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the International AAAI Conference on Weblogs and Social Media* (2011), ICWSM '11, pp. 50–57.
- [7] BLACKBURN, J., AND KWAK, H. STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games. In *Proceedings of the 23rd international conference on World Wide Web* (2014), WWW.
- [8] CORREA, D., SILVA, L. A., MONDAL, M., BENEVENUTO, F., AND GUMMADI, K. P. The Many Shades of Anonymity: Characterizing Anonymous Social Media Content. In *Proceedings of The 9th International AAAI Conference on Weblogs and Social Media (ICWSM'15)* (Oxford, UK, May 2015).
- [9] GORDEEV, D. Automatic verbal aggression detection for Russian and American imageboards. *CoRR abs/1604.06648* (2016).
- [10] HOSSEINMARDI, H., GHASEMIANLANGROODI, A., HAN, R., LV, Q., AND MISHRA, S. Analyzing Negative User Behavior in a Semi-anonymous Social Network. *CoRR abs/1404.3839* (2014).
- [11] INGRAM, M. Here's Why You Shouldn't Trust Those Online Polls That Say Trump Won. <http://for.tn/2dk74pG>, 2016.
- [12] JOHNSON, A., AND HELSEL, P. 4chan Murder Suspect David Kalac Surrenders to Police. <http://nbcnews.to/2dHNcuO>, 2016.
- [13] LEAGUE, A.-D. Pepe the Frog. <http://www.adl.org/combating-hate/hate-on-display/c/pepe-the-frog.html>, 2016.

- [14] NOBATA, C., TETREAUULT, J., THOMAS, A., MEHDAD, Y., AND CHANG, Y. Abusive Language Detection in Online User Content. In *WWW* (2016), pp. 145–153.
- [15] ONWUZURIKE, L., AND DE CRISTOFARO, E. Experimental Analysis of Popular Smartphone Apps Offering Anonymity, Ephemerality, and End-to-End Encryption. In *UEOP* (2016).
- [16] PEDDINTI, S. T., KOROLOVA, A., BURSZTEIN, E., AND SAMPEMANE, G. Cloak and Swagger: Understanding data sensitivity through the lens of user anonymity. In *S&P* (2014).
- [17] POTAPOVA, R., AND GORDEEV, D. Determination of the Internet Anonymity Influence on the Level of Aggression and Usage of Obscene Lexis. *ArXiv e-prints* (Oct. 2015).
- [18] POTAPOVA, R., AND GORDEEV, D. Detecting state of aggression in sentences using CNN. *CoRR abs/1604.06650* (2016).
- [19] RAINIE, L. The state of privacy in post-Snowden America. <http://pewrsr.ch/2d216bm>, 2016.
- [20] RIVERS, C. M., AND LEWIS, B. L. Ethical research standards in a world of big data. *F1000Research* (2014).
- [21] ROESNER, F., GILL, B. T., AND KOHNO, T. Sex, Lies, or Kittens? Investigating the use of Snapchat’s self-destructing messages. In *Financial Cryptography and Data Security*. 2014.
- [22] SORGATZ, R. Macroanonymous Is The New Microfamous, 2009. <http://fimoculous.com/archive/post-5738.cfm>.
- [23] STEIN, J. How Trolls Are Ruining the Internet. <http://ti.me/2bzZa9y>, 2016.
- [24] THE WEEK. The rise of the alt-right. <http://theweek.com/articles/651929/rise-altright>, 2016.
- [25] WAKEFIELD, J. Net overload sparks digital detox for millions of Britons. <http://bbc.in/2ayPMU8>, 2016.

9. RARE PEPES

In this Section we display some of our rare Pepe collection.



Figure 28: A somewhat rare, modern Pepe, which much like the Bayeux Tapestry records the historic rise of /pol/.



Figure 30: A (French?) Pepe wearing a beret, smoking a cigarette, and playing an accordion.

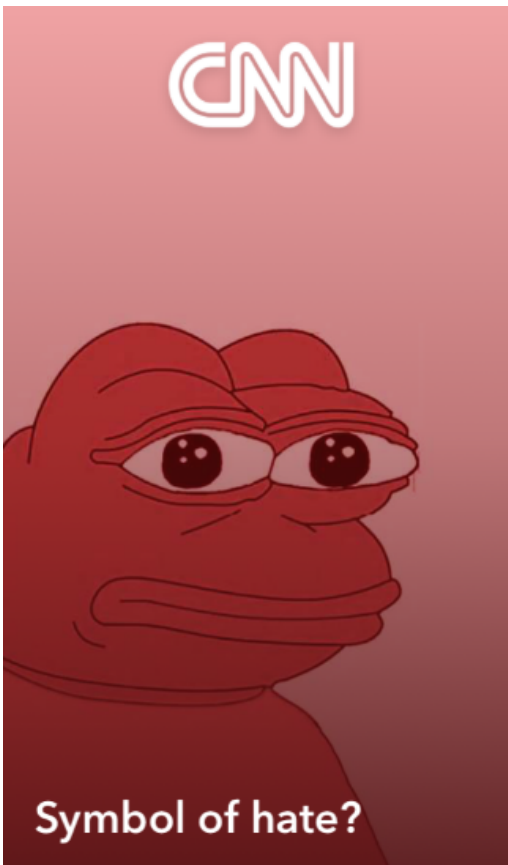


Figure 29: An extremely common Pepe commissioned by CNN to commemorate Pepe's recognition as a hate symbol.



Figure 31: An (unfortunately) ultra rare Pepe eating a delicious Publix Deli Sub Sandwich.



Figure 32: An ironic Pepe depiction of Hillary Clinton.



Figure 34: What we believe to be a Pepe re-interpretation of Goya's "Saturn Devouring His Son."



Figure 33: A Pepe Julian Assange dangling a USB full of Democratic National Convention secrets.



Figure 35: A very comfy Pepe.



Figure 36: A mischievous witch Pepe.



Figure 37: The now “iconic” Trump Pepe.