

Discord: The Next Big Thing? A First Look at Discord’s Growth, Servers, and Users

Ethan Chiu
Wellesley High School
17chiue@gmail.com

Eni Mustafaraj
Computer Science Department
Wellesley College
emustafa@wellesley.edu

Abstract—Discord, a voice chat and instant messaging platform that went live in March 2015, started as a meeting place for video games enthusiasts. It expanded quickly, without drawing mainstream attention, until shortly after the “Unite the Right rally” in Charlottesville where the New York Times reported Discord becoming “favorite chat app” of alt-right and it had been used for advertising and gathering support for the rally. Since apologizing and removing these hate groups, Discord has grown to 87 million users. To date, there has been no in-depth research on this platform yet. How do we study the growth and characteristics of a platform that supports decentralized anonymous communities? Our study provides the first in-depth look at Discord’s platform, growth, servers, and users.

I. INTRODUCTION

With the growth of video game use among teens, Discord has created a platform in which people can communicate quickly and effectively. Due to its privacy, accessibility, and features, Discord has reached beyond gamers and has been a powerful tool for communication. To our surprise, despite the recent popularity in the press and its more than 87 million user base, we found very little evidence of researchers examining Discord during the time period since its establishment in March 2015. Specifically, we were only able to find an article by a librarian [1] who summarizes Discord among many other social media platforms, and a few articles about Discord in the context of gaming [2], [3]. In the meantime, the latest magazine profile about the platform [4] has the CEO of the company boosting about a user base of 87 million. We don’t have a way to verify this number independently. Furthermore, it is not clear whether the number refers to unique individuals, or account names, given that users are allowed to have multiple identities. In any case, its magnitude is impressive.

Unlike public platforms like Twitter and Facebook, the availability of easy-to-use APIs and the public nature (most user profiles are public) make these platforms very appealing to researchers. But the Web is big and many users are looking for new platforms to engage. If these platforms don’t make it easy for us to collect data, should we walk away, or should we think creatively about the Web as an ecosystem where interactions leave traces that can be collected and followed?

In our paper, we tackle this question and provide a basis for (1) identifying a platform headed towards significant mass appeal gives researchers the opportunity to engage early in descriptive and prescriptive research; (2) testing and implementing new data gathering techniques for studying emerging

platforms is useful to support research that goes beyond “Twitter research”; (4) how users communicate on a decentralized platform.

Our contributions in this paper are three-fold: (1) We provide a comprehensive description of Discord and its features that is currently missing in the research literature of social media and online communities; (2) we present a series of ways that serve to reconstruct the breakout path of Discord, without relying on any data from the Discord platform itself, but by only making use of its footprints on other platforms; (3) we analyze how users interact on a Discord server; (4) we analyze messages collected from two Discord servers for presence of hate speech, and discuss the difficulties of automatically recognizing it. The structure of our paper follows these three goals exactly: describing the platform’s anatomy; collecting and analyzing its footprints on the Web ecosystem; understanding how users interact on a Discord server; and detecting the presence of hate speech.

II. RELATED RESEARCH

A. Research on Early Platforms

Early papers on new social media platforms like Twitter [5] and Flickr [6] have used the platforms’ public APIs and crawled the platform. [6] crawled Flickr using their API once every day to gather links and user information, gathering 9.7 million links and 950,000 users. Since they had links, users, and user groups that captured all the data within their time range collected, they used their [7] methods to create Barabasi-Albert (BA) models and power-law degree distributions to chart the growth of Flickr. Similarly, [5] collected 1.3 million Twitter messages over the course of two months using Twitter’s API. Twitter’s API provided user ids, messages, networks between users, and timestamps, allowing them to graph Twitter’s users growth rate and user intention. They found that the number of users nearly doubled, from 3.4 million to 6 million users.

The Discord platform is by design decentralized, allowing users to organize in countless servers, private and public. The number and details of all Discord servers are unknown (unless Discord decides to publish them, but that seems unlikely at the moment), so it’s impossible to replicate an approach like [6], to gather all data from Discord each day for a certain time span. Similarly, although there is a Discord API for creating bots, there is no API to gather the number of users and messages. Lacking direct access to a representative source of data from

Discord, we resorted to footprints of the platforms in the Web ecosystem to understand its growth over time.

B. Research on hate and internet culture

There is a strong body of research on the culture of anonymous platforms. For example, [8] using 20.7 million Whisper messages, found that anonymity led to more aggression and more inclination to truly express their needs and desires. [9] gathered 8 million 4chan messages from a variety of boards every day, allowing them to see what posts are archived and collect all messages and user activity. They found that Americans dominated the conversations on /pol/ (the politically incorrect board) and 4chan's users performed raids on other platforms like YouTube using signal processing techniques. [10] gathered 5.7 million 4chan posts and found that ephemerality and deletions play a strong role in shaping internet culture, anonymity was a strong part of community identity. In 2015, Reddit implemented bans on multiple community it deemed hateful such as r/fatpeoplehate and r/CoonTown. [11] researched on what happened to Reddit after those groups were banned through 670 million comments and posts on Reddit in 2015, which were publicly available. They discovered that banning these communities reduced hate speech on the platform by more than 80 percent since it prevented trolls and hate mongers from communicating with each other in safe spaces. The fact that Discord combines voice chat with text messages might lead to underestimating the real extent of hate speech, given that some of it might occur in real-time voice chat and is not captured for posterity.

C. Research on community growth

There has been significant research on what factors contribute to community growth. [12] uses 3,000 Ning communities and [13] collected WikiProjects since March 2013, finding that groups with a high short term growth rate are more likely to continue to grow long term. [12] also found that users connected previously to a group were more likely to join a highly clustered group and that structural features predict longevity of a community the best. Similarly, [14] found that communities that overlapped in the same niche lead to high-activity levels.

These findings combined with our findings about possible audience overlap between Reddit and Discord or YouTube and Discord, suggest several avenues for further research that we discuss in the final section.

D. Research on sustaining communities

There has also been significant research on maintaining a successful community. [15] argues that small groups and commitment from leaders of online communities fostered more group outcomes. To encourage commitment and contributions to communities, they suggest that clustering people based on identity and creating named groups fosters identity-based commitment to the community. [16] found that sustaining a community relies on the features of communication's activity and that the growth of social platforms relies on providing

benefits to members presently and balancing the negatives and positives of size and communication. Similarly, [17] through surveys on Everything2.com found that the most important factor for a user to stay in a community was a sense of belonging. [18] analyzes all posts from 283 Reddit communities and found that user retention was the most crucial part for community maintenance. [19] found on Facebook, that an initially engaged new user who receives comments on his or her photo will increase his or her posting activities. Our section "Anatomy of the Discord Platform" about Discord's features and structure shows that Discord has implemented several of these findings. The use of bots to maintain community seems particularly effective and begs for more research.

III. ANATOMY OF THE DISCORD PLATFORM

Discord is a new platform that combines features from several existing online platforms. The fact that it has been successful in attracting many users in a small amount of time indicates that it has a low entry barrier. In the rest of this section, we dissect the components of the platform and describe what makes it so appealing to different groups of users, including the ones who engage in hate speech.

A. Servers

Discord provides both IM (Instant Messaging) and VoIP (Voice over IP) chat. It is usually described as a hybrid between the Slack team messaging application and the Skype video chat platform.

The main components of the Discord platform are the common spaces known as servers, where users gather together to chat, primarily in real-time. Each server has a unique URL; it can be public or private; and Discord supports the creation of hundred of thousands of them. Everyone can create a server for free and then send invite links to their friends or virtually anyone interested to discussing a topic to join the server. There is no limit in the number of invitations or the number of users who can join a server.

B. Users and User Roles

A user who creates a server is known as an "owner". An "owner" can assign the "admin" role to other users that they invite to the server. Admins can then send invites to more users and assign other roles to them. Another common role is that of a moderator, or "mod" for short. Roles have privileges attached to them. Privileges are actions that a user can perform in the platform, such as ban members, create instant invites, change a nickname, manage emojis, manage webhooks, and many more. Often, admins use "bots" to perform various tasks such as assigning roles to new users automatically. Bots are a common and widespread presence in Discord.

C. The User Interface

Figure 1 shows a typical server. The left part of the interface in a server is dedicated to text and voice channels, which users can join, based on their roles. A channel's name starts with '#', and one can see them like hashtag streams in Twitter,

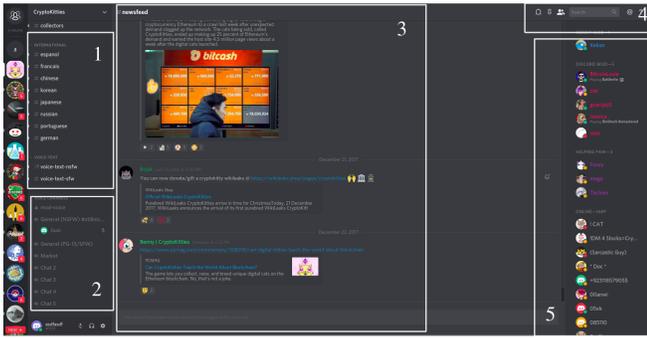


Fig. 1: Example of a server with a channel section highlighted. It contains (1) text channels, (2) voice channels, (3) text channel, (4) options to mute the highlighted channel, view pinned messages, and search messages, and (5) current active users separated by roles

except you have to be within the server to participate in the channel. The middle section contains a typical chat stream of messages. Users can also upload files such as images, GIFs, and videos, which then appear in the stream. The right section displays the list of active users, grouped by roles. Users can at any time start private conversations with other users. There are also two common types of bots in Discord servers: responsive and Webhook bots[20]. For example, users can interact with a responsive music Discord bot by inputting the message !play in a channel and users can view news articles from a webhook bot that gets articles from a site.

IV. DISCORD DEPICTION BY ONLINE TECH NEWS SOURCES

Since going live in 2015, Discord has seen a very fast growth, with the latest registered users count amounting to 87 million [4]. This seems even more impressive when compared to Twitter, which after three years had only 54 million registered users [21]. However, from the very beginning Twitter appealed to a more mainstream and professional audience, attracting tech journalists and researchers who publicized its features for real-time news gathering early on [5]. Contrarily, Discord positioned itself as a platform for supporting gamers and young users, and this narrow focus kept it outside the interests of mainstream media and researchers.

Date	Users (millions)	Source
2016-01-26	2.9	[22]
2016-07-08	11	[23]
2016-12-09	25	[24]
2017-05-16	45	[25]
2017-12-08	87	[4]

TABLE I: Total number of registered users on Discord according to various news sites

Table I summarizes the data that Discord reported to VentureBeat, Gamasutra, TechCrunch, Mashable, and Rolling Stone. The last four rows show an exponential growth with factor 2 almost every 6 months.

V. FOLLOWING THE FOOTPRINTS OF DISCORD THROUGH WEB ARCHIVES

One way to establish the pattern of growth for an online platform is to look for past news articles, as we showed in the previous section. Is it possible to consult other sources to independently verify the claimed growth? We decided to try the Internet Archive’s Wayback Machine¹, a platform that has been archiving the web since 1996 and allows the public to access or upload snapshots of different web pages. In the recent years, the platform has seen wider use especially by social scientists [26]. However, this archiving is not perfect, and studies have shown that new websites take long to be archived for the first time. Additionally, accessing the archived data automatically poses many challenges too, and one is advised to verify the results manually.

In the following, we show the results we received by tracking the number of Discords servers and Discords impressions on social platforms over time.

A. Tracking Discord Servers:

Discord doesn’t publicly reveal the number of the servers in its platform, however, several websites maintain community-aggregated lists of public servers. One can easily find many of such websites by googling for “discord servers”. We found several, discord.me, discordlist.me, discord-server.com and others. After identifying several such websites, we tried to find their archives on Wayback Machine and encountered various difficulties. Many of these websites are formatted through pagination (numbered pages: discord.me/servers/1, discord.me/servers/2, etc. However, it appears that the Wayback Machine doesn’t follow such links. Thus, although there are currently 111 snapshots of the landing page for the website discord.me, the biggest aggregator of Discord servers, the snapshots are not archives of the entire site, which we will need to calculate the total number of servers. We were lucky with another website, discordlist.me, which indicates the total number of the listed servers in its landing page. However, this website has only four snapshots available, showing another limitation of the archiving (its delay and infrequency). The earliest snapshot is from June 2017, listing 4048 servers and the latest on from January 2018, listing 14018 servers, an increase with factor 3.4. Lists from other websites either had overall a small number of servers, or their archiving had errors that made it hard to observe the trend.

B. Tracking Discord’s Social Media followers:

The Wayback Machine also archives pages from social media platforms, and due to their popularity, these platforms are archived more frequently. We manually gathered the number of likes, subscribers, or followers on Discords official space on Facebook (discordapp), Reddit (r/discordapp), and Twitter (@discordapp). The collected numbers are depicted in the chart in Figure 2 using a log scale. There is an upward trend

¹<https://archive.org/web/>

among all platforms as time elapses. It is interesting to notice the differences: Reddit page was the first to be captured, the Twitter account has most followers (571K as of January 2018), and the Facebook page has been created much later. The growth in Twitter users is particularly impressive, going from 177K in Jan 2017 to 571K in Jan 2018. This growth led us to look for Discord traces in Twitter directly, which we show in the next section.

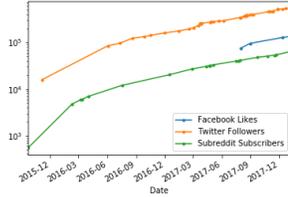


Fig. 2: Social followers of Discord’s main pages across Facebook, Twitter, and Reddit (log scale)

In conclusion of this section, it is worth emphasizing that the Wayback Machine does allow us to discover evidence that supports the argument for exponential growth of Discord, but several technical challenges exist and there should be a better way to search and analyze data from the Wayback Machine.

VI. DISCORD FOOTPRINTS IN OTHER PLATFORMS

The Discord platform has its presence in social networks, such as Twitter and Reddit. However, it’s not equally easy to find out how big this presence is and how it has grown over time. As usually, the platform that makes it easier to access data is Twitter, and we will show the results below. Additionally, we used a special searching feature on Google, to assess the current footprint of Discord on other social media and networking platforms.

A. Twitter interactions from and with Discord:

Discord has a Twitter account, @discordapp, which is used to engage with followers on Twitter. We used a twitterscraper library² to collect all tweets sent by Discord @discordapp, as well as all tweets directed at @discordapp. We were able to collect 196,657 tweets sent by @discordapp and 562,143 tweets that are interactions between @discordapp and various Twitter users.

Figure 3 represents a time series of the above-mentioned tweets organized by month, since the creation of the discordapp Twitter account. Some of the anomalies in the graph (such as the bars for Sep and Oct 2016) might be an artifact of the scrapper script we used. Despite the obvious errors in data collection, there is a clear trend of increase in the Twitter interaction between Discord and its followers, especially since June 2017.

Meanwhile, Figure 4 represents a time series of the various counts of retweets, likes, and replies towards the most popular tweets by Discord in every month.

²<https://github.com/taspinar/twitterscraper>

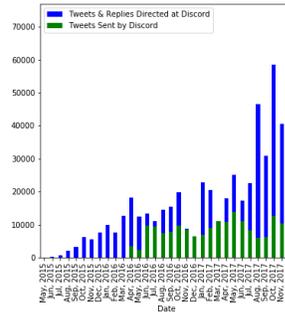


Fig. 3: Discord sent tweets & # of tweets and replies involving Discord.

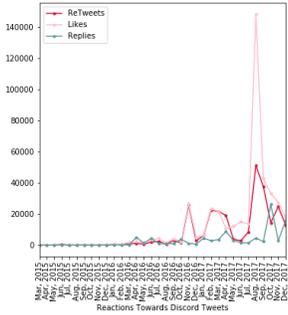


Fig. 4: Reactions for Discord’s most popular tweets.

The highest peak in the graph corresponds to the tweet³ with which Discord notified the shutdown of the alt-right related servers after the events in Charlottesville. Other peaks corresponds to promotional tweets for give-aways.

B. Google’s Trends, Correlate, and Hits

Google collects different kinds of data, for example, the users interests in the form of the volume of their searches over time, as well as user-generated content that is stored by different websites. We looked for footprints of discord in both these data sources.

To follow users interest in Discord, we turned to Google Trends⁴, a tool provided by Google to track the volume of a query over time. Figure 5 shows the trend graph for the query “discord”. Before March 2015 (Discord’s release date), it is nearly a straight line indicating, constant volume for the English word “discord”. After Discord’s release, the searches have been increasing steadily over time, showing the same pattern of almost doubling every year.

To evaluate how often other websites contain Discord-related content, we searched Google using the filter “site:siteURL” together with our search term. The filter restricts to results that only come from that site.

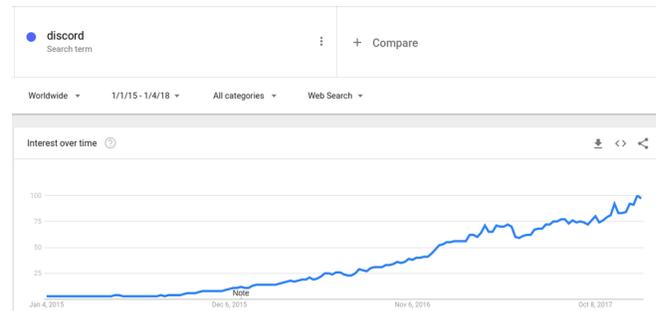


Fig. 5: Google Trends chart for the query “discord”.

Table II reflects the total number of search results for different sites, sorted in descending order. The search term

³<https://twitter.com/discordapp/status/897170310348263426?lang=en>

⁴<https://trends.google.com/trends/>

Search Term	Number of Hits
discord	33,800,000
discord site:youtube.com	11,600,000
discord site:reddit.com	6,380,000
discord site:twitter.com	1,060,000
discord site:facebook.com	608,000
discord site:instagram.com	20,600
discord site:4chan.org	19,200

TABLE II: Google search terms and the number of results for the query “discord”. Retrieved in January 2018.

“discord” is ambiguous because it might relate to (1) the English word for conflict; (2) an Anime character; (3) the Discord platform. However, we assume that the other two word senses are uniformly distributed in all these platforms, leaving us with an ordinal arrangement of platforms as shown in Table II. Ignoring the hit number for Google, which aggregates the hits over all websites on the Web, the table provides us with a new picture of where conversation about Discord happens the most: YouTube and Reddit. This can be explained with the overlap of the primary audience of Discord, gamers, with those of these two platforms. YouTube is used to host videos of games or of people playing games, so gamers use it frequently. Reddit is also a platform where gamers organize in many subreddits. In fact, a common form of advertising for Discord servers is to share their invitation links on topically affiliated subreddits.

Search Term	Number of Hits
whatsapp	1,190,000,000
whatsapp site:facebook.com	8,420,000
whatsapp site:youtube.com	6,120,000
whatsapp site:twitter.com	2,710,000
whatsapp site:instagram.com	1,630,000
whatsapp site:reddit.com	155,000
whatsapp site:4chan.org	None

TABLE III: Google search terms and the number of results for the query “whatsapp” (retrieved on January 2018). Notice that the ranking of websites here is very different from that in Table II, for example, see Reddit’s position and hits.

Correlation	Search Term
.9913	bypass google
.9910	2017 kia
.9905	google lock
.9902	panzoid
.9896	coupe 2017
.9878	afc urgent care
.9873	what is discord
.9871	discord music
.9857	2017 toyota tundra
.9853	2017 ford explorer

TABLE IV: Google search term “discord” and its top ten correlations. Correlation = $\text{corr}(\text{“discord”}, \text{Search Term})$ Retrieved in April 2018.

We also looked at possible correlations between the time-series of Discord searches versus other topic using Google

Correlate⁵ which allowed us to see the highest correlated time series of search terms with our own search term. Since Discord was public released in May 2015, we searched correlations of the search term “discord”. Table IV shows the top 10 correlated time-series. The top correlation “bypass google” seems to indicate that Discord’s anonymity and features have been enticing for hackers and programmers to congregate on Discord’s platform.

C. Tracking through Bot Analysis

In addition to public Discord server lists, there are also public lists of bots that Discord users can use to choose and install bots on their server. The most popular one of them all is Carbonitex⁶ who has its own bot which many servers use to gain statistics about their site. Thus, we used this site to gain insights on how many servers have bots and how many users the most popular servers.

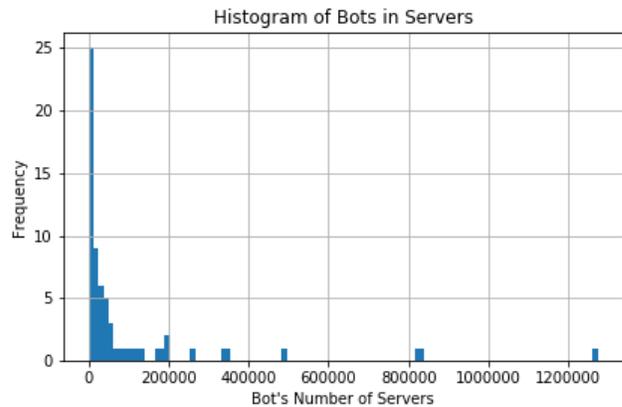


Fig. 6: Histogram of bots’ number of servers

From carbonitex’s bots page, we were able to scrape data on 65 public bots and the number of users who installed each bot. We found our sample to have a mean of 102,107 users and range to be between 1 to 127,3357. Figure 6, we can get a fair estimate on how many servers use bots and get a sense on how many servers there really are. It’s skewed to the left

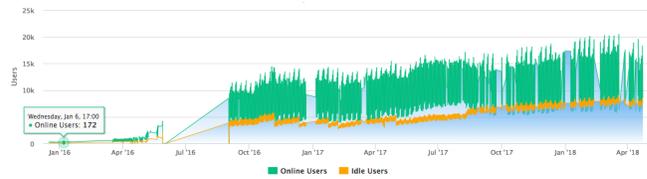


Fig. 7: Discord server Overwatch online vs. idle users over time (hourly)

Since Carbonitex is installed on many popular Discord servers, we can also see how many users are in each server and how active they are⁷. In Figure 7, we show a graph of

⁵<https://www.google.com/trends/correlate>

⁶<https://www.carbonitex.net/>

⁷<https://www.carbonitex.net/discord/servers>

how active the most popular server, Overwatch, which has over 107,000 members as of April 2018. This figure shows that the Overwatch server has steadily increased member and maintained an active group of users over time.

D. Tracking through online web trackers

There are also websites which track the traffic among other statistics for websites. As of April 2018, according to the Alexa ranking website⁸, Discord ranks 146 globally. As a point of comparison, Facebook ranks 3, Twitter ranks 13, and Reddit ranks 6 globally. Moreover, Alexa calculates that 40.7% of users come from the United States, 5.7% from the United Kingdom, 4.5% from Canada, and 3.7% from Japan. It's impossible to know how Alexa calculates these demographics and rankings but this information provides an important basis for comparison in our own data.

In conclusion of this section, it is important to point out that there are a variety of ways in which one can reconstruct the growth path of one platform. A common factor however seem to be the traces the users leave on various social networking platform: their interaction with the Discord account on Twitter, their searches over time on Google, and their mentions of Discord servers on platforms such as YouTube and Reddit. The number of Google hits within a site seems to be a useful indicator that can lead to further analysis. To see the usefulness of such numbers, in Table III we summarize the same search phrases, but this time for the query "whatsapp" (a popular VoIP chat application). The results show differences in audience overlap.

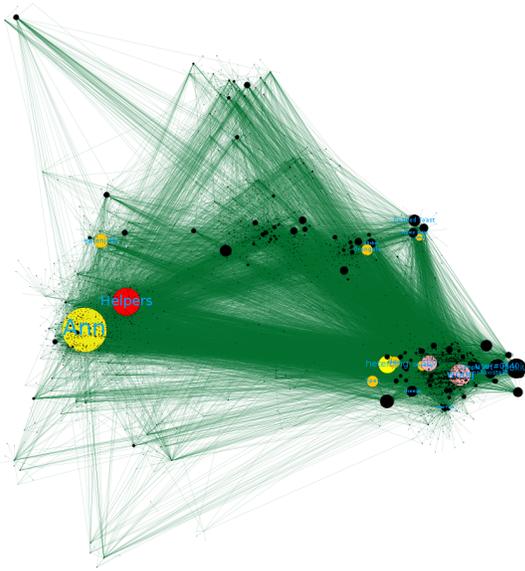


Fig. 8: Social Network of the "Mathematics" Discord server. (yellow = founder, red = group role, orange = moderator, pink = admin, black = user (no role))

⁸<https://www.alexa.com/siteinfo/discordapp.com>

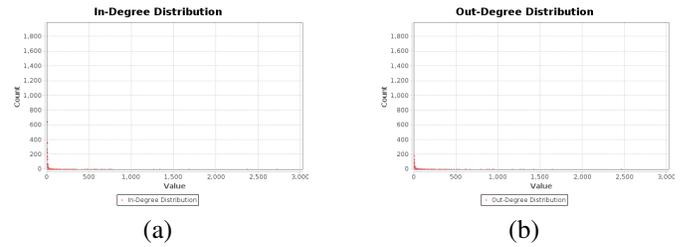


Fig. 9: Weighted a) in-degree (user/group mentioned) and b) out-degree (mentioner) distributions of the "Mathematics" Discord server.

VII. USER INTERACTIONS ON DISCORD

We focused on one Discord server to analyze. We chose the most popular educational community we could find which was the "Mathematics"⁹ server, a community of 4661 users (as of April 10, 2018) who discuss and help one another out on math concepts, to look at how users communicate with each others and bots. For this data-set, we scraped 1,408,917 messages using the Discord Scraper¹⁰ from when the "Mathematics" server was created on January 13, 2017 to April 10, 2018. In addition to scraping these messages, we joined the server on January 29, 2018 and observed qualitatively how users interacted with one another in various channels.

To get a better sense of how users interact on the servers, we wrote a script which parsed through each message that we collected and collected the users and groups mentioned in each message. Then, we constructed a two lists: one which contains the sender ID and one which contains the user/group mentioned ID. We counted 52994 users and groups mentioned.

Using this data produced by our script, we then created Figure 8 (N=4721, E=23380) using the OpenOrd algorithm [27] which shows groups vertices based on average-link clustering and forced-direct layout. Figures 9 show the weighted in-degree (user/group mentioned) and out-degree (mentioner) distribution. For Figure 8, we chose to manually color and label the top 20 nodes which had the highest weighted in-degree (# of mentions) for clarity. Since users can have multiple roles, the color indicates the highest role that the node (user or group) has.

This graph shows some interesting clusters such as on the left side which contains the most mentioned user: "Ann". She was one of the founders and was communicating quite frequently with other users to help them understand mathematical contacts. Close to this node is the "Helper" role where users would call out to get math help. The bottom right cluster contains mostly "moderators" and "admins" who maintain the server and makes sure rules of each server is enforced.

For bots in the servers, they were some of the most vocal in the network which makes sense. The "MathBot" had 21st highest weighted out-degree value (413) while "Tatsumaki" bot had the 26th weighted out-degree value (363). In our

⁹<https://discord.me/math>

¹⁰<https://dht.chylex.com/>

observations, we found that the "MathBot", which helps search Math definitions and format math text, was utilized when user(s) ask for homework help from a "Helper". Moreover, it's interesting to see the "Tatsumaki" bot being interacted with considering which rewards fake points for users who are active in a Discord server. As opposed to other platforms like Twitter and Facebook, bots on Discord interact directly with other users, augmenting user interactions in the process.

Overall, our methods of constructing a social network of the "Mathematics" Discord server shows a way to utilize prior methods of constructing social networks for centralized platforms onto a totally new, decentralized platform. Moreover, we describe how Discord's unique functionalities such as its bot integration and user roles impact a Discord server.

VIII. IS HATE-SPEECH TOLERATED IN DISCORD?

News articles in several outlets have pointed the finger at Discord, claiming that it allows hate speech to flourish [28], [29]. Discord defends itself by stating that communities self-regulate themselves, and that when notified of abusive behavior, Discord suspends servers or users, as in the aftermath of Charlottesville's rally.

Identifying hate speech in online communities is challenging. Additionally, different communities have a different level of tolerance for it. We describe here both a qualitative and quantitative analysis that compared two similar communities who had a role in the Charlottesville rally.

A. Data Collection

To collect messages on Discord servers, we used the Discord History Tracker¹¹.

Prior to the shutdown of alt-right due to the Charlottesville's rally, we captured 2,700 messages from the "formal_discussion" channel of the alt-right server, covering the period Aug 1-6, 2017. Technical issues with the scraper prevented us from collecting more data, or data closer to the date of the Charlottesville rally.

Similarly, we captured 2,979,089 messages from all text chat channels of the server Centipede Central from Oct 10, 2016 to Aug 15, 2017. Centipede Central is the offshot of the subreddit "r/The_Donald", which similarly to its parent, is frequented by Trump supporters [30]. The subreddit "r/The_Donald" has also been labeled as "hate-mongering" in some media [31]. Shortly after Charlottesville's rally, Centipede Central was raided by their own admins, because of their dislike of the server's owner [32]. As a result, many messages were deleted. Our third batch of messages belongs again to Centipede (after the raid): 26,126 messages covering the time interval Sept 28, 2017 to Jan 15, 2018.

Our goal was to compare the presence of hate speech in messages from the alt-right server and the two datasets of Centipede Central. We decided to keep the Centipede data in two different batches, to check if post-Charlottesville, there was a change in the messages, given that the administrators of

the servers expressed to reporters their worry of being banned from Discord [33].

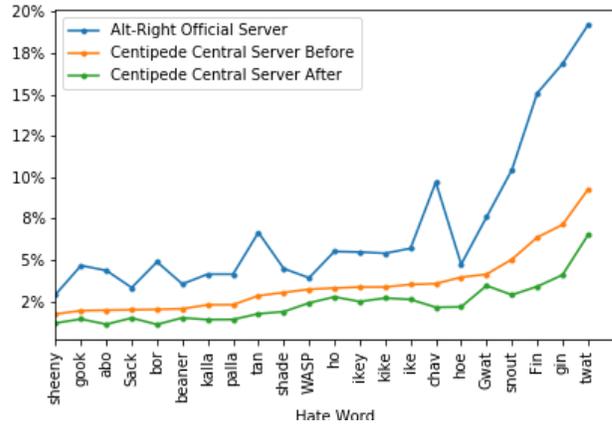


Fig. 10: Stacked line graphs of the percentage top 22 shared common hate words in messages in the alt-right server and Centipede Central servers (during and after Charlottesville)

B. Quantitative Analysis:

Our approach for detecting messages that contain hate speech mimics that used in [9] for analyzing 4Chan posts. That is, we also used the hatebase¹² API to get the top 1000 hate words used in the United States (the API has a country parameter) and the NLTK framework¹³ to create a dictionary of similar variants for each hate word.

By analyzing every message for the presence of at least one hate word, we found that 12.26% of the messages in the alt-right server contained them, which was twice the percentage found for messages in Centipede Central before the rally: 6.55%. There was a decrease of the use of hate-speech in Centipede Central after the rally, with only 4.26% of messages containing hate words.

Then, similarly to [9], we focused on the most frequent hate words. Figure 10 shows a graph of the top 22 shared hate words and their prevalence in the three datasets.

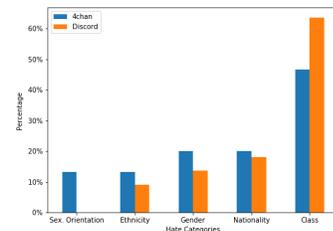


Fig. 11: Side by side comparison between the percentage of top hate words from 4chan and percentage of top hate words from Discord, grouped by category.

The hatebase API provides categories for hate words, and we grouped the hate words in this paper and those from [9] in

¹¹<https://dht.chylex.com/>

¹²www.hatebase.org

¹³<http://www.nltk.org>

one common graph, see Figure 11. The more pronounced existence of class-related hate words in our dataset is interesting, but a topic for future research.

IX. DISCUSSION AND FUTURE RESEARCH

We believe to have provided an interesting and necessary first-look at a fast-growing online platform that became an unexpected player in the complex political atmosphere of 2017, joining the ranks of the established platforms such as Twitter (with its Russian bots problem) or Facebook (with its fake news problem). The lack of previous research on Discord made it necessary for us to explain the structure and features of the platform, as well as explore several avenues for establishing how the platform was able to grow quickly. Our insights may serve as a starting point for further research in some promising directions, listed below.

1) *The role of bots*:: The Discord community is heavily invested in bots, both as guardians of communities and as facilitators of communication. While other communities such as Reddit and Wikipedia also make use of bots, the Discord bots seem far more sophisticated and nuanced, thus, a systematic analysis of their features, and more importantly their acceptance within the communities is a worthy pursuit.

2) *Overlapping communities*:: Evidence from Google hits results suggests that Discord users might be also active on YouTube and Reddit. While challenging, it seems relevant to find ways to track how communities interact across multiple platforms and how such interactions fuel growth of new platforms, as we suspect happened with Discord.

3) *Experimental Evidence*:: Observational approaches start to break down once APIs for accessing data are unavailable. While we explored several alternatives to the lack of comprehensive data from the platform, they will not be able to explain **why** Discord grew so quickly. In addition to user surveys, one needs to setup servers and channels to become part of the Discord ecosystem and to study from within what motivates users and what tools are efficient in keeping abuse and hate under control.

REFERENCES

- [1] K. E. Anderson, "Getting acquainted with social networks and apps: Social media in 2017," *Library Hi Tech News*, vol. 34, no. 10, pp. 1–6, 2017.
- [2] H. Korkeila and J. Hamari, "The relationship between players gaming orientation and avatars capital: a study in final fantasy xiv," in *Proc. of the 51st Hawaii ICSS*.
- [3] A. Waugh, "Teen nerdfighters: An exploration of engagement in a complex fan community," *Proceedings of the Association for Information Science and Technology*, vol. 54, no. 1, pp. 821–823, 2017.
- [4] B. Crecente. (2017) Discord: 87m users, nintendo switch wishes and dealing with alt-right.
- [5] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 56–65.
- [6] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhat-tacharjee, "Growth of the flickr social network," in *Proc. of 1st workshop on Online Social Networks*, 2008, pp. 25–30.
- [7] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhat-tacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM*. ACM, 2007, pp. 29–42.
- [8] D. Correa, L. A. Silva, M. Mondal, F. Benevenuto, and K. P. Gummadi, "The many shades of anonymity: Characterizing anonymous social media content," in *ICWSM*, pp. 71–80.
- [9] G. E. Hine, J. Onaolapo, E. De Cristofaro, N. Kourtellis, I. Leontiadis, R. Samaras, G. Stringhini, and J. Blackburn, "Kek, cucks, and god emperor trump: A measurement study of 4chan's politically incorrect forum and its effects on the web," in *ICWSM*, pp. 92–101.
- [10] "4chan and/b: An analysis of anonymity and ephemerality in a large online community."
- [11] E. Chandrasekharan, U. Pavalanathan, A. Srinivasan, A. Glynn, J. Eisenstein, and E. Gilbert, "You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, p. 31, 2017.
- [12] S. R. Kairam, D. J. Wang, and J. Leskovec, "The life and death of online groups: Predicting group growth and longevity," in *Proc. of the 5th ACM WSDM*, 2012, pp. 673–682.
- [13] J. Solomon and R. Wash, "Critical mass of what? exploring community growth in wikiprojects," in *ICWSM*.
- [14] H. Zhu, J. Chen, T. Matthews, A. Pal, H. Badenes, and R. E. Kraut, "Selecting an effective niche: an ecological view of the success of online communities," in *Proc. of the SIGCHI Conference*, 2014, pp. 301–310.
- [15] R. E. Kraut, P. Resnick, S. Kiesler, M. Burke, Y. Chen, N. Kittur, J. Konstan, Y. Ren, and J. Riedl, *Building successful online communities: Evidence-based social design*, 2012.
- [16] B. S. Butler, "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Information systems research*, vol. 12, no. 4, pp. 346–362, 2001.
- [17] C. Lampe, R. Wash, A. Velasquez, and E. Ozkaya, "Motivations to participate in online communities," in *Proc. of the SIGCHI conference*, 2010, pp. 1927–1936.
- [18] J. Zhang, W. L. Hamilton, C. Danescu-Niculescu-Mizil, D. Jurafsky, and J. Leskovec, "Community identity and user engagement in a multi-community landscape," *arXiv preprint arXiv:1705.09665*.
- [19] M. Burke, C. Marlow, and T. Lento, "Feed me: motivating newcomer contribution in social network sites," in *Proc. of the SIGCHI conference*, 2009, pp. 945–954.
- [20] Discord. (2018) Channels resource discord's api.
- [21] L. Wolfe. (2017) Twitter user statistics 2008 through 2017.
- [22] D. Takahashi. (2016) Discord raises \$20m for voice comm app for multiplayer mobile games.
- [23] B. Francis. (2016) Game chat app discord crosses 11 million registered users.
- [24] S. Perez. (2016) Gamer chat app discord hits 25 million users, can now be used in developers own games.
- [25] A. Rosenberg. (2017) Discord's slack-meets-skype service is growing quickly among gamers.
- [26] S. K. Arora, Y. Li, J. Youtie, and P. Shapira, "Using the wayback machine to mine websites in the social sciences: a methodological resource," *Journal of the Association for Information Science and Technology*, vol. 67, no. 8, pp. 1904–1915, 2016.
- [27] S. Martin, W. M. Brown, R. Klavans, and K. W. Boyack, "Openord: an open-source toolbox for large graph layout," in *Visualization and Data Analysis 2011*, vol. 7868. International Society for Optics and Photonics, 2011, p. 786806.
- [28] B. Menegou. (2017) How a video game chat client became the webs new cesspool of abuse. [Online]. Available: <https://gizmodo.com/how-a-video-game-chat-client-became-the-web-s-new-cessp-1792039566>
- [29] K. Roose. (2017) This was the alt-rights favorite chat app. then came charlottesville.
- [30] J. Koebler. (2017) The secret chatrooms where donald trump memes are born. [Online]. Available: https://motherboard.vice.com/en_us/article/qk994b/the-secret-chatrooms-where-donald-trump-memes-are-born
- [31] A. Romano. (2017) Reddit just banned one of its most toxic forums. but it wont touch the_donald. [Online]. Available: <https://www.vox.com/culture/2017/11/13/16624688/reddit-bans-incels-the-donald-controversy>
- [32] J. Koebler. (2017) Internets most popular donald trump chat room destroyed in anime-related coup. [Online]. Available: https://motherboard.vice.com/en_us/article/a3kq94/what-happened-to-centipede-central-discord-trump-memes
- [33] S. Cole. (2017) Pro-trump discord server 'centipede central' says it's being monitored. [Online]. Available: https://motherboard.vice.com/en_us/article/zmmxe/pro-trump-discord-server-centipede-central-says-its-being-monitored