

THE UNIVERSITY OF CHICAGO

**True or False Incels:
an analysis of the Incels.is forum**

By

Zachary Hinds

July 2022

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Brooke Luetgert
Preceptor: Brianne Painia

1 Introduction

In 2014 in Isla Vista, California, Elliot Rodger murdered six people and injured twelve more (Yan et al., 2014). Rodger considered himself a member of the Involuntary Celibate, or Incel, Community, an online community of men. Specifically, the Incel identity and community is founded upon a shared hatred towards women and society based on the perception that Incels are unjustly excluded from romantic and sexual relationships with women. Since his horrific attack, Rodger has been elevated to martyrdom within the Incel Community and both his manifesto and the Incel Community as a whole have gone on to inspire several more killers to commit similar acts of violence, such as the even deadlier Toronto van attack in April of 2018 (Austen and Stack, 2018; Witt, 2020). Among the Incels who commit these attacks (as well as the community which inspired and often supports their actions still), such extreme violence is the justified response to the injustices faced by Incels within society (O'Malley et al., 2020).

The current research on Incels has presented an in-depth psychological profile. The Incels can be understood through the traits and narratives which drive many individuals to isolate themselves from everything but the community and can lead a dangerous few to go so far as to turn to violence (O'Donnell, 2021; Scaptura and Boyle, 2020). Furthermore, the Incels present an intensely isolated, enraged, and reactive masculinity which is driven to further intensity both in reaction to perceived threats to their own masculinity as well as to the social pressures on men to value themselves primarily through the lens of dating and their sexual/romantic relationships (Cosma and Gurevich, 2020; Hoffman et al., 2020; Scaptura and Boyle, 2020). Particularly, the isolation of Incels has proven to be a difficult obstacle in studying Incels, especially given Incels have developed intense hostility towards researchers or other figures of authority who stand to potentially delegitimize their world (Ging, 2019).

Beyond the study of individuals, research which has expanded the scope out to the community level has also highlighted the specific aspects of the Incel community which set it apart from traditional extremist spaces. One primary difference is the purely digital nature

of the Incel community. The Incel community is, in part, a reaction to the stigma (real or perceived) of their social failures. This reaction pushes for social isolation, and without access to a digital platform the Incels would be unable to coalesce into any coherent, shared community around their Incel identity. Furthermore, the Incels formed within an online context characterized by anti-women and anti-feminist narratives referred to collectively as the *manosphere* (Ging, 2019). Considering this context, it is important to note the unique characteristics of Incel community movements and of the community’s larger structure. As an extremist group online, they have been subjected to a variety of forced digital migrations as their communities on larger social media spaces like Reddit (with the likes of /r/Braincels or /r/Incels) have been banned or suppressed. These migrations have a real effect both on the population size of these communities as well as on the intensity of extremist rhetoric (Horta Ribeiro et al., 2021).

Incels have been identified in connection with the rising threat of misogynistic violence (Center, 2021). This status indicates the importance of developing a deeper insight into the role that the Incel forums and similar digital spaces play in the intensifying of these sorts of identities. Social media and online forums have enabled the creation of these sorts of ideology-focused groups which can inspire a sense of community separate from that of real-world interactions and can spur the creation of extreme subcultures which lead individuals to justify violent or antisocial behavior (Holt, 2007). This is especially apparent in the case of Incels both given the real cases of violence influenced by the community and through the explicit vilification of women and legitimization of violence and revenge within Incel discussions (Holt, 2007; O’Malley et al., 2020; Witt, 2020).

Addressing the threat of the Incels has proven to be a difficult task. Efforts to de-platform Incels through widespread bans have been able to temporarily stymie the community’s growth. Unfortunately, these bans are accompanied by an increase in intensity as the more extreme elements of the community are distilled through platform migrations (Horta Ribeiro et al., 2021). Along the same lines, efforts towards prevention and treatment

have focused primarily on the individuals within these communities (Van Brunt and Taylor, 2021). In order to address Incel groups and communities like them, it is becoming increasingly apparent that more information is needed. Identifying the emergent structures from community-specific features, as well as developing a clearer understanding of the general structures of extremism as it exists in the digital space will allow for the better recognition of these sorts of groups in their infancy. Furthermore, developing a better understanding will provide a clearer path for both prevention and the development of recourse for tackling the violent potential of these sorts of communities (Van Brunt and Taylor, 2021; Witt, 2020).

The Incel community exists in several digital spaces, with evidence of Incel activity present on Reddit (though their communities are often subjected to website administrator intervention), 4Chan, 8Chan, and similar sites. Outside of these general spaces, Incels share a lot in common with Pick-Up Artists ¹, with many Pick-Up Artist communities being considered something of an introduction or gateway into the Incel community (Bratich and Banet-Weiser, 2019; Ging, 2019). Online extremist groups like Incels have also become a point of concern for newer digital platforms such as the group-chat app Discord, which in April of 2021 banned over 2,000 extremist communities from its platform (Allyn, 2021). As the digital space continues to grow and evolve, the methods of studying, identifying, and addressing these sorts of communities must evolve to keep up.

1.1 Current Aims

This thesis seeks to build upon the understanding of the social structure and narratives of the Incel community and its members, as well as identify the mechanisms which lead Incels to become entrenched in their Incel identities, driving them further into social isolation and even potentially pushing them towards committing acts of horrific violence. Additionally, this thesis seeks to demonstrate the methodological value that the combined use of network

¹Pick-Up Artists, also called PUAs, refer to the community of men who seek out or teach techniques that men can use to successfully “pick-up” women with the intention of having sex with those women or to establish a romantic relationship with them.

and language analysis has in the research of online hate groups like the Incels.

2 Data

For the analysis of the Incel Community in this thesis, specific focus was placed on the popular Incel forum, *Incels.is*. This forum was selected primarily due to its size and its popularity among Incels. The site boasts over 300,000 threads and nearly 17,000 members. Additionally, the forum was also selected due to its rich potential as a corpus. Forums provide excellent insight into the community as users create threads starting discussions or asking questions to the broader population of forum users. Furthermore, the comments are ordered sequentially in threads, allowing for the analysis of discussions as comparable to conversations with the added benefit of being of easy access for data collection through web crawling.

Instead of pulling data from the live forum, an archived version of the forum was retrieved from the Internet Archive. Using an archived version of the forum allowed for greater confidence in web scraping methods. Development of the custom web-scraping was unaffected by any potential changes to the website’s layout or code. Furthermore, it allowed for a static copy of the forum which is accessible for future re-compiling of data from the same version of the forum in future.

In this archived version, the forum was divided into four sub-forums. These sub-forums divide the forum’s discussions into narrower topic of discussion within the broader topic of the forum itself. The sub-forums and their descriptions are outlined in Table 1. Within each sub-forum are lists of threads, each with the thread title, its author, its post-date, the time of the most recent comment on that thread, number of replies, number of views, and in some cases a forum label, a form of user-driven categorization of forum threads. Threads can be filtered based on thread author, activity, label, etc. At the top of the first page, sectioned off from the other threads, are posts pinned by forum admins or moderators. Each thread

consists first of an initial parent comment created by the thread’s author and then each subsequent comment is appended to the thread.

Table 1: Sub-forum Descriptions

Sub-forum	Intended Topic
Inceldom Discussion	Discussion about Incel identity, the <i>blackpill</i> ²
Offtopic	General discussion, not explicitly about Incel ideology
Meta & Feedback	Users offer suggestions and feedback, meta-discussion about the forum website
Ban Appeals	Banned users can make a thread or comments requesting (and often arguing) for the removal of their account’s ban

Data was pulled from the archived site through a custom designed web crawler, specifically created to crawl through the pages of forum threads in each of the sub-forums, following the links to the threads to download the threads’ pages. These downloaded pages were then processed using BeautifulSoup in Python to pull specific aspects of the threads’ HTML (Richardson, 2013). For each thread the following features were collected: thread title, thread label, thread creation timestamp, and thread author’s username. Then, for each thread, the following features were collected from the comment replies in the thread: comment text, commenter username, commenter user title³, and comment timestamp.

Usernames were not considered in the analysis and instead each username being stripped from the threads and replaced with a sequential identifier of “User N ” for each member of the Incel forum included in the corpus. Comments were often dense with emotes/emojis and embedded media such as images, YouTube videos, tweets from Twitter, links to Reddit posts, and so on. These elements were removed from comments, leaving just the text content of each message.

The corpus used in this analysis consisted of 80,762 comments distributed across 2,011 threads coming from 2,410 individual users. Each of the sub-forums were represented in

²blackpill defined on Table 2.

³“User title” refers to a short word or phrase displayed below a user’s username, this is either determined by an administrator (such as reading “Banned” when a user is banned) or is user generated, varies user to user.

the corpus, though the exact representation varied based on the popularity of individual threads. The comments analyzed were created within the date range of November 8th, 2017 to August 30th, 2021⁴, though the threads themselves were between November 8th, 2017 and December 12th, 2018. These threads and comments consisted primarily of text with some threads containing embedded links, images, emotes, etc.

To prepare the corpus for analysis, all links and non-text content were removed, and the text was normalized, stemmed, and tokenized using the Natural Language ToolKit (NLTK) in Python (Bird et al., 2009). Texts were tokenized by splitting the lowercase raw text into a collection of single-word tokens using NLTK’s word tokenizer. These tokens were then further processed using NLTK’s PorterStemmer, simplifying many words down to their root (such as by dropping -ing, -er, etc.) in order to facilitate more accurate computational analysis of the text. Additionally, commonplace words were dropped using NLTK’s built in list of English stopwords and usernames were also removed from the tokens.

Incels’ language can be difficult to decipher for readers unfamiliar with the community due to the frequent use of Incel-specific slang terms throughout the forum, both in comments and in thread labels (Waśniewska, 2020). To address this specific obstacle, several Incel slang terms were identified and defined through the reference of the Incel-maintained wiki project, *Incels.wiki*. This wiki project contains a glossary page which served as the primary reference in the development of the slang reference document used for this analysis. Certainly, there are potential concerns in the use of a collaborative wiki project due to the potential for misleading or bad-faith changes to the wiki which could pollute any data collected from it. After careful consideration, this was not considered a relevant issue as the definitions provided within this glossary matched with the definitions interpreted through the contexts in which the terms were used. Furthermore, *Incels.wiki* is intended by its maintainers as a means of recruitment and for spreading their ideological framework. Because of this objective, the

⁴Though the original archive was created November 2018, due to the way that the Internet Archive API works, sometimes posts which were not captured in an earlier snapshot will be linked instead. Removing these “future” comments had no effect on any of the data, so they were left in.

Table 2: Top Slang by Count with Definitions

Term	Count	Definition
incel	3956	An involuntary celibate man.
foid	2058	A derogatory term for a female.
chad	1243	A physically attractive male, typically white.
cope	1125	Adopting an untrue belief to avoid confronting reality.
fakecel	1111	A person who falsely claims to be an Incel.
cuck	1093	A man with an unfaithful girlfriend/wife, also used as an insult, to refer to someone with (perceived) progressive political beliefs or weak and ineffectual men.
normie	818	An average everyday boring person. Sometimes used to refer to people that are “average” in looks in contrast to Chad and Incel.
blackpill	738	A fatalistic/depressed version of red pill, particularly one focused on external solutions as opposed to individual ones or on the importance of physical attractiveness in driving men’s dating success.
jfl	642	Just fucking lol.
inceldom	588	The condition of being Incel.

wiki maintainers pay careful attention to keeping its pages accurate to their message and the image they wish to portray. It is for this same reason that wiki was not further integrated into this analysis as it did not provide insight into the community itself, only the specific claims and “evidence” that Incels reference to defend their worldview. Table 2 contains some of the most frequently used slang within the Incel corpus alongside a definition based on those provided by *Incels.wiki*.

Identifying the demographics of the Incels using the comments from *Incels.is* presents significant obstacles, primarily as users usually do not openly identify themselves beyond their username. There may be the occasional passing mention of characteristics, often those that they deem to be the source of their unattractiveness, but nothing concrete. It can be inferred that all Incels on the forum are men, as it is specifically outlined in their forum rules that women are automatically banned without question. More specifically, the forum moderators are also generally quick to ban users who are deemed to be “fakecels” (a fake Incel), a broad category of the men who have had sex or a relationship with a woman but

falsely claim to be Incels⁵. The common ages of Incels are unclear, generally Incels appear to be adolescents given their discussions often make mention of classes, school, and other high school or college-related topics, but previous research has also demonstrated that there exists a large population of older members as well (Jaki et al., 2019).

Likewise, it is also difficult to determine racial demographics of the Incels. Some reports have portrayed Incels as a majority white group, but this is likely a misrepresentation. While the racism and anti-semitism within the Incel community points to a connection between Incel ideology and white supremacy, Ging (2019) presents clear evidence that there is a large non-white population within the Incels. Users often declare that white men have an easier time getting women, and there are a large number of Incel slang terms used to refer to various raced Incels (Jaki et al., 2019). It is important to not overlook this aspect of the Incels, especially when attempting to identify the factors which drive members to commit themselves to the Incel identity and worldview.

3 Methods

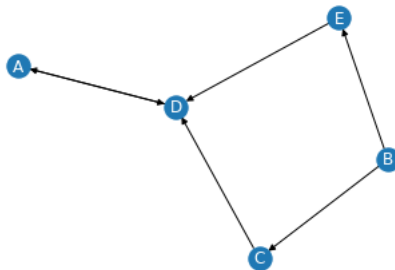
The Incel forum represents a dense collection of features which is difficult to penetrate without an equally in-depth analysis. To facilitate this analysis, this research made use of a combined approach of quantitative and qualitative methods loosely based on the mixed methods approach used by Jaki et al. (2019). Where their approach focused on a linguistic understanding of the Incel forum, this research focused its structural and narrative characteristics. Quantitative methods used a combination of network analysis, sentiment analysis, and topic modeling. Network analysis provided a broad overview of the structure of the forum while sentiment analysis and topic modeling were used to outline the community narratives. The quantitative methods then informed the specific direction of the qualitative analysis of individual posts and comments, highlighting the discussions around narratives

⁵Fakecel is also extended to would-be Incel men who are deemed overly sympathetic towards women and/or “normies.”

and structures identified with the quantitative analysis. Together these methods allowed for the targeted analysis of the structure and user interactions which took place within the forum.

3.1 Network Analysis

The thread-comment structure of the Incel forum was ideal for mapping user interactions through the construction of a network graph. The network graph was constructed through the use of the Python library NetworkX (Hagberg et al., 2008). Individual users were represented as nodes within the graph and interactions were modeled through the edges connecting these nodes. These edges were also given direction which indicated when the source node commented on a post created by the target node. Given Users A, B, C, D, E, the following network would be constructed based on user interactions.



Additionally, the edges were given a weight, w , calculated based on the number of directed interactions, n , in a linear fashion.

$$w = n_{(\text{source} \rightarrow \text{target})}$$

This accounted for instances where a user might repeatedly respond to the threads created by the same user and allowed for a more complete representation of the importance of user nodes. In total, the network consisted of 2,409 user nodes connected with 18,365 edges. Interactions where the source and target user were the same were ignored as they did not offer additional information and made the network harder to understand.

After creating the nodes and edges using NetworkX, they were then imported into Gephi in order to facilitate exploratory analysis of the complete network graph (Bastian et al., 2009). Figure 1 was generated using Gephi and shows the full network graph of the Incel corpus with nodes and node labels scaled proportionally to the nodes' degrees. The positions of nodes were determined using a force-directed model which clustered nodes algorithmically (Hu, 2005). To take advantage of Gephi's functionality for graphing dynamic networks, nodes and edges were given a timestamp attribute of their origin within the data set. In essence, nodes were created when a user first posted a thread or a comment and edges were created when a source user first interacted with a target user.

Both Gephi and NetworkX were also used to generate descriptive statistics for the network. Nodes' degrees were calculated both weighted and non-weighted and both their in-bound and out-bound degrees were also considered. User nodes' Betweenness Centrality was calculated using the methods offered by NetworkX and was normalized using the built-in function. Betweenness Centrality was calculated both for the graph when considering its directionality and when that directionality was ignored.

This approach specifically limited analysis to that of user interactions in the form of comments on users' threads. By limiting the network construction in this way, it ignored any potential for cross interactions between users who commented within the same thread. This was decided to avoid creating an even more complex network, which might have obfuscated the specific dichotomy of thread creation and comment creation in the community's structure. Most of the Incel users in the network **never** created a thread of their own, instead only interacting within others' threads. The significance of this would likely be lost in the sheer complexity of a network that accounted for all cross-interactions. Furthermore, a network with that depth would likely range in the hundreds of thousand of edges, if not more, and would have significantly limited the ability to adequately interpret or understand the network.

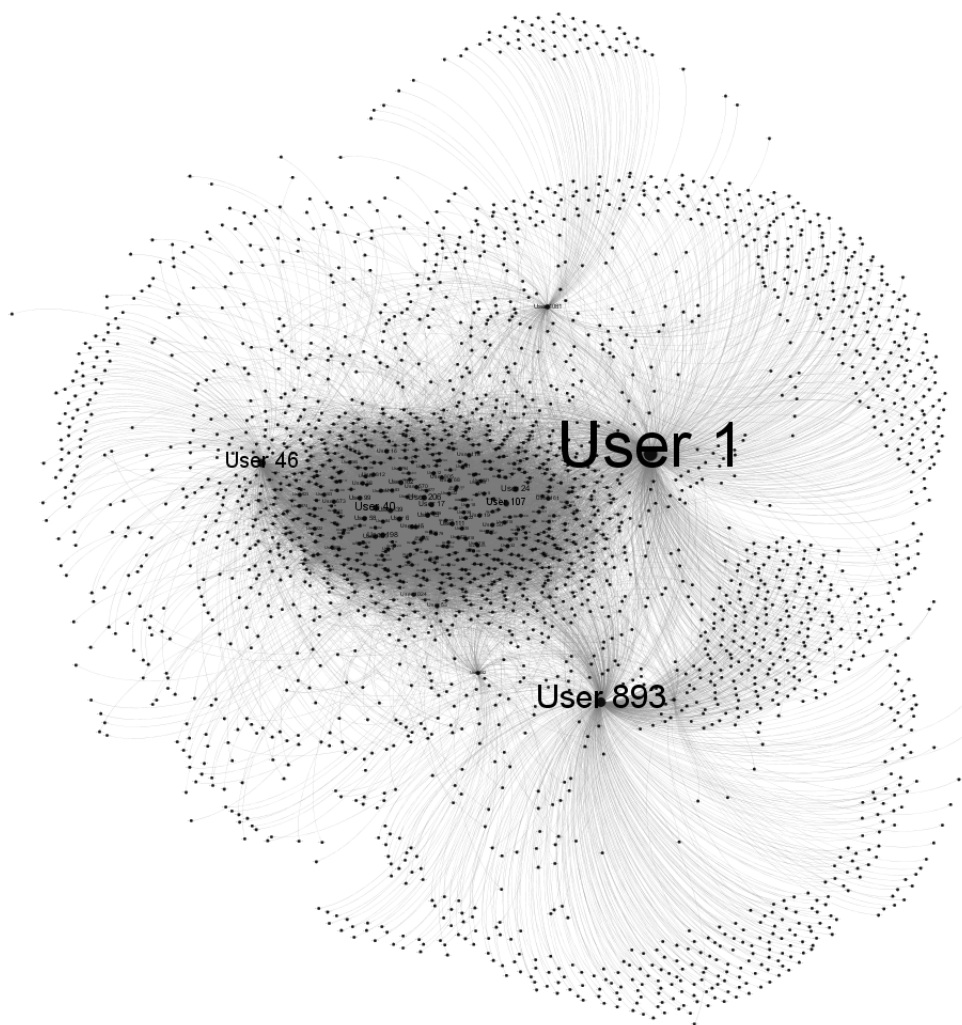


Figure 1: Complete Incel User Network

3.2 Sentiment Analysis

Interpreting the ways Incels construct their narratives and normative orders through their discussions is difficult, especially given the sheer size of the community. Computational analysis provides a potential solution through the use of sentiment analysis to construct a baseline understanding of the generalized user and comment sentiment. For each comment in the corpus, a sentiment score array was generated through the use of NLTK’s implementation of the VADER sentiment analysis tool (Bird et al., 2009; Hutto and Gilbert, 2014). Using VADER with NLTK generated a score array based on input text. Each score array had four

attributes: Negative, Neutral, Positive, and Compound. Negative, Neutral, and Positive refer to a ratio of sentiment-tagged words within the analyzed texts, whereas Compound also integrates additional rules that weight words in the input text differently. Table 3 presents an example phrase and its generated scores.

Table 3: Example VADER Sentiment Analysis

“get some blackpilled chad who hate foids like zyzz and that can be done multiple times.”	Negative	0.381
	Neutral	0.619
	Positive	0.0
	Compound	-0.5719

Comments were split into parent comments and child comments. A parent comment was the comment which starts a thread combined with the text of a thread’s title. Child comments, then, were the responding comments to that parent comment. Split this way, each comment was given a sentiment array scoring its Positive, Neutral, Negative, and Compound sentiment. Additionally, a simplified classifier was created, identifying sentiment as Positive, Neutral, or Negative based on the Compound Sentiment Score. Sentiment scores of ≥ 0.05 were considered Positive, scores < 0.05 and > -0.05 were considered Neutral and scores ≤ -0.05 were considered Negative.

The sentiment scores for each threads’ child comments were then average to create a Mean Sentiment Score for each sentiment attribute (Negative, Neutral, Positive, and Compound). The resulting 2-score pairs were then used for statistical analysis.

There were limits in the use of both NLTK and VADER given the nature of the language that Incels use. As mentioned previously, a lot of Incel slang is specific to the community and would not have appeared within the training data for most natural language tools. With VADER, terms that are not scored within its model are scored as Neutral as a default. This lead to a significant over weighting of comments towards Neutral which may not have reflected the actual sentiment of the comment. Furthermore, Incels often use sarcasm in their comments, which is difficult for current natural language tools to interpret properly. Together these two factors significantly limited the application of sentiment analysis. This

research would have been improved if these aspects had been accounted for, but there are not equivalent tools currently available to accomplish that. Other sentiment analysis models face similar issues and VADER was chosen as the best option.

3.3 Topic Modeling

In order to identify specific themes of discussion within the Incel forum, a topic model was constructed using the Gensim Python library (Blei and Lafferty, 2006). Data was prepared for the topic model by using the stemmed and sanitized tokens for each comment to a term frequency-inverse document frequency vectorizer from the Scikit-learn Python library (Pedregosa et al., 2011). This vectorizer was then used to further remove superfluous words in the corpus and create a reduced list of tokens for each comment. This reduced list was then used to create a dictionary and bag of words with Gensim.

The created bag of words and dictionary were then used to generate a Latent Dirichlet Allocation (LDA) topic model with 20 topics. This number of topics was selected in order to adequately capture the wide array of topics across the four Sub-forums. Additional models were generated with a different number of topics based on the number of sub-forums or the number of Incel normative orders as described by O'Malley (O'Malley et al., 2020). The additional models created contained overly broad topics which were difficult to make inferences from.

The 20 topics were then visualized using the pyLDAvis Python library. Figure 2 depicts the intertopic distances created and Figure 3 highlights the top terms by salience (Chuang et al., 2012).

The model was also used to determine the probability weights for each topic for each comment. The tokens of each comment were transformed into a bag of words then inputted into the LDA model, outputting the probability for each topic per comment. Additionally, for each user, a topic score was calculating by using the same process but on a collection of tokens from all of their comments combined. Together, these allowed for the comparison of

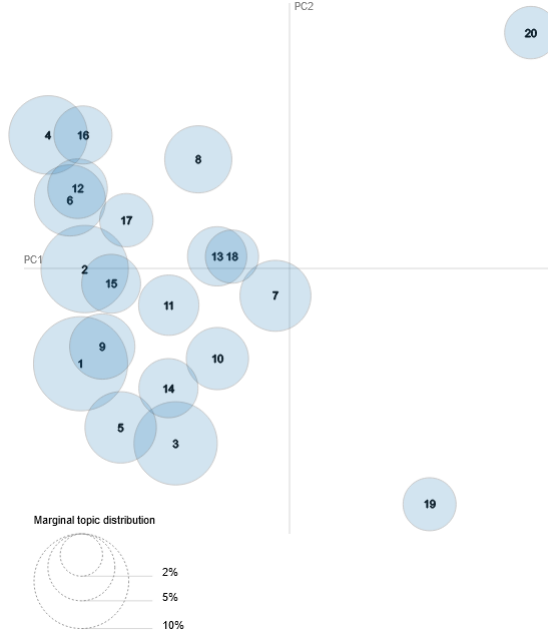


Figure 2: Intertopic Distance Map

topic weights among users and among threads.

A key limitation of the bag of words approach is the lack of consideration for grammar or context. There is no indication of the exact context in which the terms of a topic are being used and how that might affect their meanings. This can lead to issues in identify the larger discussion underneath each topic. Additionally, this issue is further intensified through the lack of accounting for tone. A term could be used in a variety of tones, be it anger, sarcasm, etc. and bag of words approaches do not have any means of differentiating them. Still, the use of bag of words allowed for a broad analysis of the corpus and still provided insights into the general aspects of the larger community discussion. For potential future analysis, larger groupings of words could be used, such as bigrams or trigrams, to provide further context, but this analysis focused on the broader insight provided by individual tokens.

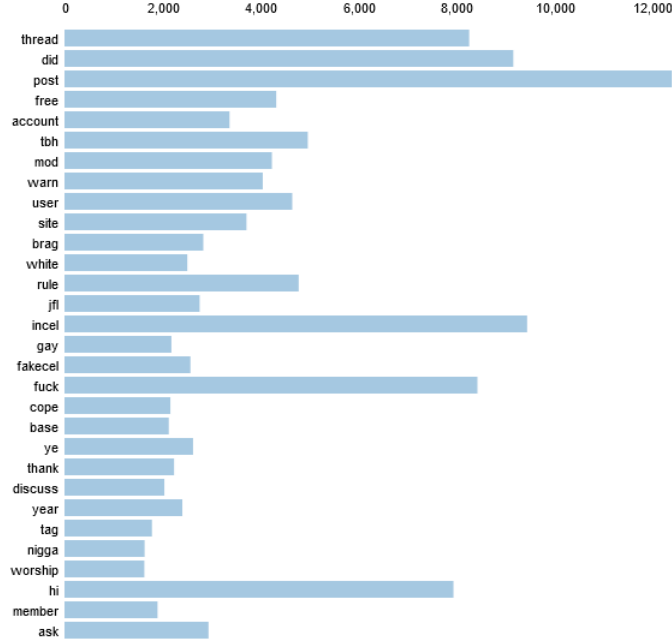


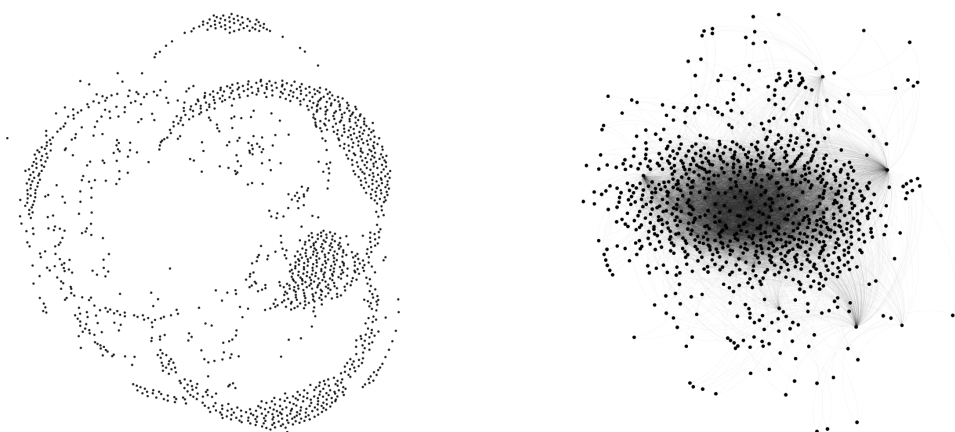
Figure 3: Top 30 Most Salient Terms

4 Findings

4.1 User Centrality

The complete network of users, as shown in Figure 1, depicts a community divided into an inner and an outer group. A large proportion of users in the network interacted with less than 3 other users over the course of the corpus. Importantly, this meant that not only is the community divided, the groups serve disparate functions. Users in the outer group are not interacting among themselves, separate from the inner community. Instead, those within the outer group concentrate around a handful of users who act as focal points along the edge of the inner community. At the same time, the inner group is largely made up of users who *do* primarily interact among themselves, with only the occasional interaction with members of the outer group. This separation is further emphasized by the Betweenness Centrality of user nodes. 59.15% of users in the Incel network had a Normal Betweenness Centrality of < 0.0001 . Thus, a majority of users exist within the outer group, largely disconnected from the inner Incel community. Figure 4 illustrates the disparity within the inner and outer

community.



(a) < 0.001 Normal betweenness centrality (b) ≥ 0.001 Normal betweenness centrality

Figure 4: Centrality Comparison

63.55% of users have a degree of ≤ 2 , indicating that they interacted with only two separate users, either by commenting on their thread or having their threads commented on. 46.33% only interacted with a single user. Furthermore, for many of these users, their edges were with the limited handful of users who act as focal points within the network. Notably, there were several sections of the outer group divisible through the specific focal-point users they interacted with. Most notably of these focal-points, were the top three users by in-bound degree and weighted in-bound degree: User 1, User 893, and User 46. In fact, these three users combined made up a significant proportion of the total weight available within the entire network. For each of these users, the question arises as to what it is exactly that places them so centrally within the greater Incel community.

User 1 was a forum moderator, thus they have several moderator-focused threads and comments. Their interactions included responding to feedback posts, discussing users' ban appeals, or generally responding as a spokesperson of the moderator team. Much of their in-bound interactions are the results of creating large "megathreads" for these purposes. Their appearance as a focal-point of the network makes sense as a result of their role within the community and is not necessarily the result of appealing to Incel narratives and establishing

themselves within the community. That said, it is still relevant that they have *formal* power over the forum. As a moderator, User 1 is able to ban/unban users, hide/remove/lock⁶ threads, and more. They also make reference to multiple older (now banned) Incel community sites in their discussions, likely placing them as a long-term member of the community.

One of the clear ways that admin users like User 1 enact their formal authority over a community is through the development and enforcement of rules. A community's rules can give clear insight into the values of that community and moderators inherently work to maintain a community built around those rules. Considering this and User 1's position, it is likely that User 1 is invested in the Incel worldview and his enforcement of the site rules works to further reinforce that worldview. This is evident with the following excerpt from a post outlining the forum's rules and terminology:

Incel (Allowed): Person who wishes to be in a romantic relationship, but is unable to despite numerous attempts and a significant amount of time trying.

Blackpilled (Conditional): Person who believes in / wants to learn about the Blackpill.

Female (Not Allowed): Banned on sight, no exceptions

Immediately, this establishes a community built on the exclusion and vilification of women. Furthermore, this also seeks to enforce the importance of determining "real" Incels and that the blackpill is also central to that determination. Further on in the rules, when outlining things which are not allowed: Immediately, these rules establish a community constructed on the exclusion and vilification of women. Furthermore, it establishes the importance of "real" Incels and centers the blackpill in that determination. Further on in the rules post, he outlines things which are not allowed on the forum including:

Cherry picking (e.g. unattractive men with girlfriends).

⁶Banning a user involves either temporarily or permanently preventing them from interacting with the forum with that account. Locking a thread means that no new replies can be made to the thread.

This preemptively seeks to dismiss any criticism and block attempts to go against the Incel worldview views all men disadvantaged in the “Sexual Market” (O’Malley et al., 2020).

In contrast to the formal position of power of User 1, User 46 did not appear to be a moderator and in fact was Banned at some point before the archive was captured. Still, User 46 created popular threads that garnered large amounts of replies and remained active for much of the corpus. Their most popular thread was one titled “I will approach⁷ 1000 girls and find a GF”:

there is no way in hell that not 1 girl in 1000 will like me I have approached 22
978 to go fellas

This thread consisted of 2,612 comments, of which 2,112 were comments created by users other than User 46. In total, 549 different users created comments in this thread over its lifespan. Based on the final comments in the thread, it was eventually closed by forum moderators, yet not before users could reminisce with comments like “I miss the glory days of this thread.” Clearly, User 46 was able to make an impact within the community and generate significant discussion. Furthermore, User 46 firmly sits within the Incel identity throughout his comment history, frequently disparaging women, boasting about catfishing⁸women, and frequently using homophobic and racist slurs.

At the time the archive was taken, User 46 had been banned from the forum. There was no official explanation from moderators for this ban that could be found within the corpus, though based on the discussion threads which mentioned User 46 following his ban, he was likely banned due to having success in his dating life and “finding a GF” and having sex. This presents an interesting case, as previous to his ban, User 46’s comments clearly supported the Incel normative order of the Sexual Market and the power women hold over dating (O’Malley et al., 2020). But if it is true that he was banned for being able to develop

⁷“Approach” is a term used frequently among Pick-Up Artist to refer to approaching random women in public spaces (such as bars, coffee shops, etc.) and using pick up lines or other Pick-Up Artist techniques to convince the women to date them or have sex with them.

⁸Catfishing refers to the practice of creating fake dating profiles online to deceive potential matches.

a romantic or sexual relationship with a woman, then his personal accounts and story would serve to delegitimize the Incel worldview. This is further supported as comments discussing him after his ban from the forum often disparaged him and declared that they had always suspected he was not a *real* Incel.

Similar to User 46, User 893 created a popular thread which maintained activity for the entire length of the corpus. In fact, User 893 created only a single thread within the forum corpus. This specific thread was the largest thread in the Offtopic sub-forum and as such was not intended for Incel ideological discussion. Instead, this thread was intended to be a place for users to post music recommendations. It consisted of a total of 2,753 replies of which only 31 comments came from User 893. In total, 485 unique users commented on the thread over its lifespan. Also unlike User 46, this thread remained active throughout the corpus and never seemed to slow or be locked by moderators.

Even though outside of his own thread User 893 was less active than User 46, he was still very much entrenched in the Incel ideology. While his comment history is not considerably large outside of his thread, his comments placed emphasis on the Incel normative orders and traits, such as his contribution to a thread titled “Chad fails suicide via shotgun to the head. Ends up disfigured. Gets first face transplant.”:

incel does the same thing no one gives a fuck and everyone makes fun of him
until he dies for real a few months later: stacy shots[sic] herself in the head for
attention: millions of views on jewtube people being sorry for her can still get
chad.

Thus, even though User 893’s own thread may not have been Incel-focused, he, like User 46, demonstrates that they are fully invested in the Incel worldview and the Incel normative orders (O’Malley et al., 2020). Unlike User 46, though, he was not eventually banned and remained active throughout the date range of the corpus.

As many users in the outer Incels never interact beyond the few focal-point users, those

focal-point users act as an intermediary between the inner and outer Incels. Outer Incels can be split based on the specific focal-point user they interact with, but within the inner community, there is no similar disconnections. There are not any outstanding subdivisions or splits within the inner Incel community, even when considering the network spans the four separate sub-forums of the Incel forum. Still, users congregate into a singular cluster of interconnections.

Another aspect of this inner and outer community is its connection to how long users have been members of the Incel community. Figure 6 plots users’ join dates against their normal betweenness centrality with users split according to user category⁹. Most noteworthy is that users among those who have a normal betweenness centrality of ≥ 0.001 , none of them joined the Incel forum after November 2018¹⁰. Additionally, even among members who fell into this power user category, those with the highest centrality primarily joined the community earlier. This demonstrates the ways in which users do not immediately join the inner community but must move after some time.

This centrality was also disconnected from the *number* of comments a user makes, as demonstrated by Figure 6. While higher centrality members did tend to have a higher number of comments on average, there were many users who joined later who did not have high centrality. This measure placed importance on the specific of who comments on a users

⁹Admins/Moderators were identified based on users with the User Title “Admincel” or “Modcel.” Power users were defined by users with ≥ 0.001 Normal Betweenness Centrality. Banned users were identified based on user with the User Title “Banned.” Regular users were all remaining users after the previous three categories were separated.

¹⁰Even when accounting for users and comments which were included from after the archive was captured, this account age disparity remained, with older accounts maintaining higher centrality.

Table 4: Focal-Point Users

	User 1	User 46	User 893
User title	“Admincel”	“Banned”	
Joined month	Nov. 2017	Jan. 2018	Feb. 2018
Total messages	1740	3612	3987
Normal betweenness centrality	0.435	0.154	0.286
Degree	1230	609	799
Weighted degree (in/out)	30480 (29995/485)	2347 (2318/29)	9673 (9654/19)

thread or whose threads that user comments on instead of on merely having a large number of comments.

The emphasis on specific users' threads further lends credence to the argument that Incels join the community on the outside, likely with less intense, less extreme ideas and slowly form connections with those already in the inner community and become further and further embedded within the inner community. Additionally, as explained previously, this further strengthens both the formal and informal authority that focal-point users have to influence the users who join the community later.

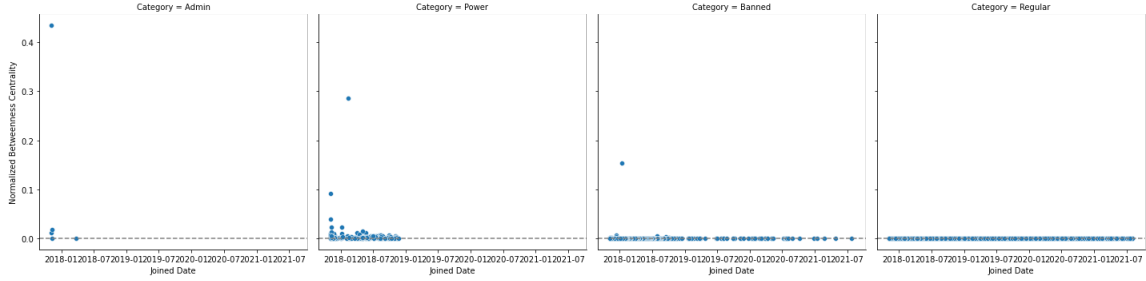


Figure 5: Relative Plots of Normalized Betweenness Centrality, Split By User Category

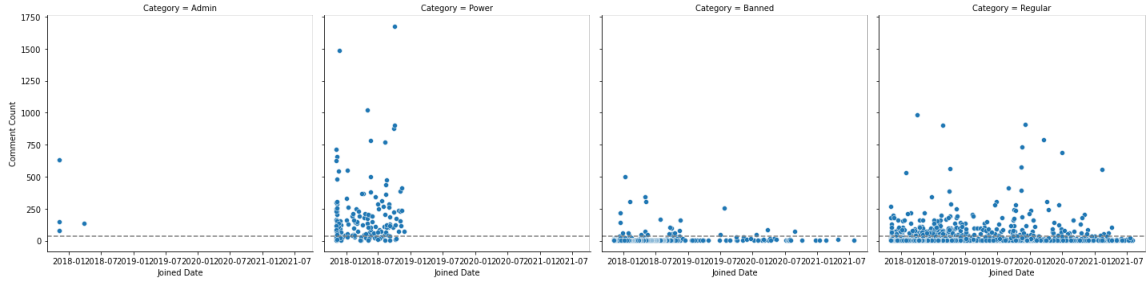


Figure 6: Relative Plots of Comments Included in Corpus, Split By User Category

4.2 Topic Modeling

Across the four sub-forums, Incel discussions covered a wide variety of topics. LDA topic modeling provided insight into the breadth of these discussions based on the tokenized corpus. For each of the 20 generated topics, a generalized descriptor was created in order to summarize the contents of that topic. These descriptor were created through a consideration

of the top terms in the topic, their collocations, and their common contexts. These descriptors were then verified by qualitatively analyzing the posts with the greatest topic weight in that specific topic.

Considering the frequency of these topics, there were a handful which appeared far more frequently than the other topics. In particular, Topic 1 (Forum Administration) out measured the other topics in the model by far. The next most common topics included Topics 6 (Sexual Marketplace), Topic 7 (Disgust), Topic 14 (Frustration), Topic 15 (Paranoia), and Topic 19 (Fantasy and Desire).

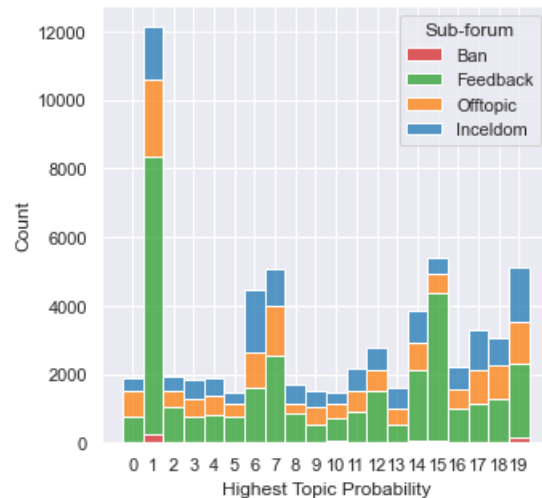


Figure 7: Topic Probability Distribution Across Comments

4.2.1 Topic 1: Forum Administration

Forum Administration was most commonly the top topic by quantity. 12,137 comments had this topic with the highest probability weight. The most common terms in this topic were focused on managing the forum, dealing with banning or unbanning users, and discussions involving moderators. The top term, in fact, was "Mod" a shorten term for forum moderators. This comment was posted by a user in the Meta & Feedback sub-forum:

it's a forum where a well-known rulebreaker got banned and the mod said to learn from his example. i'm not even saying i agree with [User] just saying that

comparing [User]’s post to an autocrat is a bit out there.

This comment came from a thread in which users were discussing the validity of another user’s ban. Discussions like this were incredibly commonplace, especially in the Meta & Feedback forum. The moderators were a central figure to these discussions, either present themselves or referred to as a looming collective. This highlights an inherent distrust of authority, even among their own members. A large contribution to this topic’s dominance was that it was the most common topic on the Meta & Feedback sub-forum. That said, even among the other three sub-forums, the topic remained dominant.

Alongside demonstrating a distrust of authority, it also highlights the specific role of moderators like User 1 in the discussions within these forums. Even outside their formal authority to control the forum, moderators also become a central figure in general discussions as users either agree or disagree with their actions on the forum.

4.2.2 Topic 6: Sexual Marketplace

Sexual Marketplace was the highest represented topic in 4,449 comments. The top words included “incel,” “foid,” and “men.” Comments within this topic often discussed things such as women’s preferences, disdain for their perceived discrimination in getting women to date them or have sex with them, and a disparaging of “normies” for not realizing the realities of men’s disadvantage against women. The following comment was posted by a user in the Inceldom Discussion sub-forum:

i really do not understand how men can not see this they just need to look for couples on the streets almost 2/3 of the couples i see the man is 2 points above the foid or even more

It was commonplace for users to discuss comparisons in the attractiveness of men and women who are visibly in relationships and it was extremely common for users to use numerical representations of attractiveness. This came in a variety of forms such as the “points” used in

this comment, though it was also common to see men or women referred to as some number out of ten.

4.2.3 Topics 7: Disgust

Disgust was the highest represented topic in 5,061 comments. Notably, among its top terms included common profanity as well as the term “volcel.” Volcel, like Incel, is a portmanteau with the word celibate, referring to “voluntary celibates” and is used by Incels to refer to people who *could* have sex or a relationship but choose (at least in the Incels’ view) not to. Commonly, this word was often used to refer to women who claim to be similarly unable to find dates. Primarily, this is because according to the Incel worldview, given women’s advantage in the sexual marketplace, these women just need to “lower their standards” to find someone.

Additionally, much of the discussions in this topic was about “fakecels,” or “Fake Incels.” These are individuals who Incels believe are falsely claiming to be an Incel often due to either that they are, in reality, sexually active or that they are “attractive enough” to find dates or sexual partners. Fakecels and volcels are targets of hatred by Incels, often considering them infiltrators in their community.

The discussions in Disgust were usually rather aggressive, making frequent use of profanity and slurs in their heated discussions, with comments such as:

lmao @[User] get fucked you fakecel faggot hopefully its a perm this time.

Here, the aggression that was directed at this “fakecel” was clear. A similar level of aggression was commonplace in a lot of the Incel discussions, but was notably present within this topic.

The notion of a fakecel and the corresponding aggression directed at them highlights a key aspect of the Incels focus on authenticity. Incels are vigilantly on the lookout for anyone who does not adequately fit their definition of Incel and are quick to single out those who go against the narratives.

4.2.4 Topic 14: Frustration

Frustration was the highest represented topic in 3,855 comments. Its top terms also included common profanity, the discussion departed from that of Disgust, more focused on frustrations the Incels felt. This frustration was often directed at the world or the women the Incels interacted with, such as this comment in the Inceldom Discussion sub-forum:

had 2 girls laugh at me yesterday on the street. what's up with people not minding their own business

Unlike the Disgust topic, this comment demonstrates that the comments within this topic were often vocally frustrated, yet did not use aggressive terms and did not appear to be as enraged in their discussions. Instead, there was a generally defeated attitude.

4.2.5 Topic 15: Paranoia

Paranoia was the highest represented topic in 5,413 comments. Its top terms were often forum related, including “thread,” “post,” “user,” and “alt.” In this context, alt refers to alternative accounts, or the other accounts a forum user might use outside of their “main” username. Alts are primarily used in order to continue to post on a forum even after receiving a ban, but can also just refer to a user’s multiple accounts that carry no such ill intentions.

These terms were best defined as paranoia given how they were used in discussions. These terms in context were often used within discussions which accused users of being outsiders using alt accounts to invade the community and avoid their bans. This comment was posted on the Meta & Feedback sub-forum:

uh so i was browsing it and stumbled across this post. how many of these infiltrators do we have? is there a way to find out who this user is?

Again there was the emphasis on frequently verifying the legitimacy of users’ Incel status and the disparaging of outsiders who were seen as infiltrating or ruining the community. As

with the Disgust topic, there existed a clear emphasis on constant verification that Incels are truly Incels and not those imposters who sought to damage their community.

4.2.6 Topic 19: Fantasy and Desire

Fantasy and Desire was the highest represented topic in 5,098 comments. Among the parent comments, this topic most frequently had the highest probability. Its top terms included “want,” “appeal,” and “sex.” Discussions in this topic were often about things the commenters wanted: their goals, desires, and fantasies. Some of this discussion was focused on forum-related goals such as appealing a ban, or wanting a new feature for the forum. Beyond those forum-focused discussions, there was clear evidence of more Incel-ideological goals or wants. For example, this back and forth took place in a Meta & Feedback sub-forum thread titled “Promoting “blackpilled” NAWALT¹¹ Youtubers Should Be Penalized:”

Commenter: “secondly you’re implying that i want to enslave women. i don’t.”

Thread Author: “women have far too many rights. this is one of the core problems. trying to appease normies by becoming more and more moderate up to a point where you’re basically a moderate cuckservative has zero appeal to me. then we do indeed have simply different goals”

This demonstrates another key element of these discussions. While there were disagreements about the extremes of what violence or goals were acceptable, there was still a tacit tolerance or even approval of those methods. In a case of “someone should do this” as opposed to “I will do this,” Incels may not outright plan their violence (though some do) their approval can, and has, led others to act in their stead (Holt, 2007; Witt, 2020).

4.2.7 Comparisons Across Users

Figure 8 shows the distribution of top topic probabilities for users, split along the same categories as Figure 5 and 6. Generally, this remained pretty consistent to what was observed

¹¹NAWALT is an acronym meaning “Not All Women Are Like That.”

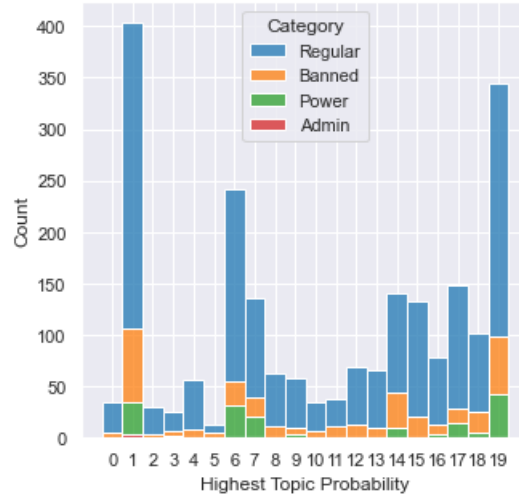


Figure 8: Topic Probability Distribution Across Users

among the comments, although the Forum Administration topic does not out measure the rest of the topics by nearly as much. Furthermore, the Fantasy and Desires topic and the Sexual Marketplace topics also had a greater presence when measured among users, while the Disgust, Frustration, and Paranoia topics had a decreased presence.

Topic 17 (Mental Illness) also became more relevant. The top terms of this topic include “iq” and “mental.” Many discussions in this topic focused on mental illness, intelligence, and users feelings of inadequacy as it relates to those things. This topic also included discussions and insults towards the intelligence of non-Incels as well. For example, the following post is from the Inceldom Discussion sub-forum:

what is your iq? iq is important for success in life so i was wondering what the average iq of incels are

In a similar manner to the Sexual Marketplace topic, Incels were heavily focused on determining their potentials for success, be it in life or in dating, based on numerical characteristics. This might be wealth, IQ, point-based attractiveness, penis length, or even the circumference of their wrists.

4.3 Reply Sentiment

Sentiment analysis was not conclusive. In order to understand the influence of parent comments over response comments through sentiment analysis, the sentiment scores were used to perform a linear regression. An Ordinary Least Squares (OLS) Regression was done through the StatsModels library for Python (Seabold and Perktold, 2010). Initially, the regression was performed using the Compound Sentiment of the parent comment as the independent variable and then the Mean Compound Sentiment of all of the thread’s response comments as the dependant variable. The resulting regression was poor. Table 5 depicts the summary of the regression results produced by StatsModels. There was not evidence of any covariance between a parent comment’s sentiment and the average sentiment of its responses.

Table 5: OLS Regression Results: Full Forum

Dependant Variable:	Mean Response Sentiment	R^2 :	0.080			
Model:	OLS	Adj. R^2 :	0.080			
Method:	Least Squares	F -Statistic:	152.8			
No. Observations:	1755	Prob(F -Statistic):	1.05e-33			
Df Residuals:	1753	Log-Likelihood:	-415.71			
Df Model:	1	AIC:	835.4			
Covariance Type:	nonrobust	BIC:	846.4			
	coef	std err	t	$P > t $	[0.025	0.975]
const	-0.0387	0.008	-5.149	0.000	-0.053	-0.024
Parent Sentiment	0.1413	0.011	12.360	0.000	0.119	0.164

Another factor considered was that the different sub-forums may require different approaches. To test this assertion, another OLS regression was performed using the same variables as the first, but this time the data was limited to only threads from the Inceldom Discussion sub-forum. The results of this regression are summarized in Table 6. Even here, there was only a marginal increase in the R^2 value, further demonstrating that there was little covariance in the sentiment of the parent comments and their responses.

This failure to establish covariance likely was the result of the limitations of sentiment analysis in Incel language as mentioned previously. These limitations undercut the accuracy

Table 6: OLS Regression Results: Inceldom Discussion

Dependant Variable:	Average Response Sentiment			R^2 :	0.107	
Model:	OLS			Adj. R^2 :	0.106	
Method:	Least Squares			F -Statistic:	57.65	
No. Observations:	481			Prob(F -Statistic):	1.65e-13	
Df Residuals:	479			Log-Likelihood:	157.11	
Df Model:	1			AIC:	-310.2	
Covariance Type:	nonrobust			BIC:	-301.9	
	coef	std err	t	$P > t $	[0.025	0.975]
const	-0.0588	0.008	-7.250	0.000	-0.075	-0.043
Parent Sentiment	0.0910	0.012	7.593	0.000	0.067	0.115

of the sentiment analysis, indicating a need for an improvement in the tools to study these sorts of communities. For future Incel sentiment analysis, it will be necessary to develop a full lexicon of Incel terminology in order to accurately score comments. On the other hand, further research is needed within natural language processing in order to account for sarcasm in analysis.

5 Discussion

The Internet provides new and expanded means for individuals to form communities around any number of topics or ideas. It has never been as easy as it is now to create a forum or website dedicated to a specific subculture. This has opened the way for the growth of communities with extremist and dangerous ideologies to create communities of like-minded individuals through which they can develop justifications to commit deadly acts of violence (Holt, 2007; O’Malley et al., 2020; Witt, 2020).

It is difficult, though, to address these online hate groups and communities and it is becoming increasingly necessary to broaden the scientific understanding of the unique ways that these hate groups, such as the Incels, manifest within digital space. Research into Incels has primarily been focused on the individual, specific characteristics of the Incels themselves. There has been limited research on the greater, community aspects of the Incels and their

formation of group-specific social norms and structures as informed by their Incel identity.

Through analysis of one of the Incels’ largest online forums, this thesis demonstrated the important relationship between the structure of the Incel community and the narratives integral to the Incel identity and worldview. Particularly, the Incels present a community organized largely into a division of inner and outer groups focused around a key handful of individual users within a narrative of harsh paranoia towards identity-purity.

The Inner and Outer paradigm presents a key aspect of the processes through which users begin on the periphery of the Incels and then move within and become embedded in the inner community. Given the harsh paranoia about the existence of so-called “fakecel” within the forum, in order to be accepted (and remain accepted) within the inner community it is highly important for Incels to engage outwardly and enthusiastically with the Incel norms and identity through their posts and comments. Incels who, in any way, go against this identity face rejection either becoming ostracized by the community or receiving forum-wide bans. Even among focal-point users, like in the case of User 46, this threat remains ever present. Even though from the beginning, User 46’s stated objective had him constantly “approaching” women, only when he succeeded in his efforts was he declared a persona non grata within the community. His “transformation” to non-Incel contradicted the Incel worldview, so the only option for the Incels, then, was to claim that he had never *actually* been an Incel in the first place and to ban him for being a “fakecel.”

Being embraced by a community like the Incels is a desirable outcome for many of the men who interact with these communities. The men who join these communities often feel isolated, excluded, and seek some form of social connection (Ging, 2019; Scaptura and Boyle, 2020). As Incels buy further and further into the Incel identities and narratives, they receive greater reinforcement to both their connection to the Incel community and to the vilification of the women and structures they view as responsible for their perceived marginalization and subjugation (O’Malley et al., 2020). Further, as users share stories of failure and disappointment in engaging with women, their masculinity is increasingly

threatened, driving them further into their belief in the Incel worldview (Scaptura and Boyle, 2020).

This investment in the Incel worldview also works as a positive feedback loop, because as Incels sink deeper and deeper into the Incel narratives, they further isolate themselves from non-Incels and while developing their social connections to other Incels. As a result, Incels increase their reliance on the Incel community as their primary social connection. Often, Incels even recognize this process. It is commonplace on the forum for Incels to request to be banned in order to restrict themselves from interacting with the community. In fact, this is so common that there are specific rules and procedures on the forum for requesting bans (and unbans in the event that they decide to return).

Additionally, the investment in being embraced by community also points to an answer to the question of how individual men come to join the Incel community. Within their circumstances and social isolation, many of the Incels also indicate some recognition of social inequality. This manifests within Incel discussions with frequent discussions of perceived racial disparities in dating, of the struggles of mental illness, and of the advantages which come with wealth. Thus, the Incel worldview provides a convenient “solution” to these problems, as well as a clear target for their frustrations and anger. Furthermore, the Incel worldview also latches onto the threats to the patriarchal order presented by Feminism and increased rights afforded to women, providing the community a focus around the strengthening and restoration of patriarchal norms and structures (Bratich and Banet-Weiser, 2019).

Finally, the drive to remain a “truecel” (and thus maintain the social connections formed within the Incel community) also explains the tacit approval of the more extreme elements of the Incels. So to avoid appearing as a “normie” or “fakecel,” users react to suggestions of physical or sexual violence without challenged. Users will do little to respond to these sort of violent discussions outside perhaps a sarcastic dismissal or claims that users must be “just joking” about the violence. Furthermore, they ignore the many comments which allude to violence as a certain eventuality or necessity in order to right the injustices they

faced (O’Malley et al., 2020). If the community standard is to accept this violent rhetoric as part of the discussion, it becomes necessary for Incels to engage with the more violent lines of thinking in order to maintain their place within the community.

5.1 Further Study

An aspect of the Incels that was not discussed within this analysis but which was apparent in the development of this research was the emphasis on the sexualization of children, specifically young girls. Research into Incels has primarily focused on the propensity for misogyny, hatred, and physical violence, yet little has been said about the acceptance of pedophilia within the Incel community. There is significant focus on virginity and women’s purity, often in the context of disgust towards women who are not virgins and those who have had multiple sexual partners. The other side of this focus of virginity is the fantasizing and idolization of the virgin woman. This fantasy is not limited to adult women, and, in fact, is primarily targeted at those who are teenage or younger, and, sometimes quite explicitly, young girls. This represents a necessary direction that Incel research should take, especially in determining if there is evidence that this approval is leading these men to act on their fantasies and harm children.

The shortcomings of the sentiment analysis performance was primarily a result of the difficulty in parsing the language of the Incels. Incel slang is difficult to decipher and filled with sarcasm and metaphor (Waśniewska, 2020). This difficulty indicates a need for further development into tools for analyzing the language of these communities. Already, there is significant work demonstrating the potential for the automatic detection of hateful language (Jaki et al., 2019). Thus, even with the limitations of current language models, there are clear indications of their strong potential in the study of language online. Better language models will also allow for a deeper understanding of the Incel identity as it manifests within these sorts of online communities, such as through the development of machine learning tools for more effective analysis of similar large-scale corpora (such as outlined in Appendix A).

De-platforming Incels has failed to prevent their growth or spread, instead the community simply finds and relocates to new digital spaces to expand their community and intensify their hatred (Horta Ribeiro et al., 2021). Even if this particular forum was removed, the Incels would merely need to create a new forum or website to facilitate their community, something which would occur almost immediately. Because of this, it is paramount that further research strive towards the development of tools and methods which can work to adequately address the threat that Incels pose. This thesis demonstrates the importance of understanding the structure and language of the Incels. Further research should continue to build upon this knowledge and hopefully provide a path towards the development of effective tools for dealing with the growing danger of the Incel community.

5.2 Beyond Incels

The methods use for this analysis can also be implemented in the study of similar online extremist groups or movements. The quantitative analysis of online communities through mapping social interactions to a social network provides a detailed overview of the structure of the community and topic modeling can provide the narratives which underscore community discussion. Together these methods provide researchers with significant insight into the structure and narratives of a community and can inform targeted qualitative analysis. These methods also can emphasize the key individuals or social focal points within a community and can indicate the necessary direction of further investigation. Furthermore, topic modeling can unveil the hidden narratives that underscore community discussion.

These methods are not limited to forum based interactions, and could also be repurposed for use within other social media spaces such as Twitter or Facebook. The insight offered by these tools provides a foundation for the development of a clear understanding of both the scope of a particular hate group online as well as provides a rich context for identifying the group-specific hate speech and narratives. Developing a concrete understanding of the language and social structure of online hate groups provides an important foundation for

further research both into the group itself and into the development of methods for addressing the danger they pose.

While there are currently limitations in the application of sentiment analysis and natural language processing in online hate communities, these tools should not be discarded. Within the study of Incels, research has already demonstrated the potential for algorithmic detection of hateful language (Jaki et al., 2019). As the tools improve, the potential scope of research into online extremism can expand and allow for greater solutions to the threat that hate groups pose.

Appendix A Machine Learning Model

Initially, for this research incorporated the development of a machine learning model to assign scores to forum text based on Incel traits established in prior research (Scaptura and Boyle, 2020). For this purpose, approximately 300 Incel forum threads were manually scored on twenty Incel traits. Each thread was scored based on a consideration of agreement with a statement developed for each trait to determine its relevancy to the text. These traits and questions are outlined in Table 7.

Originally, this research included an attempt to develop a machine learning model based around the assignment of Incel traits (Scaptura and Boyle, 2020). To achieve this, post parent comments were to be hand scored for each of the twenty outlined Incel traits. These scores were determined through a rating system which considered the text of a comment against a statement developed for each of the traits. The scorers then rated each of these statements as Strongly Disagree, Disagree, Slightly Disagree, Neutral, Slightly Agree, Agree, Strongly Agree. The traits and statements are outlined in Table 7. Agreeance was then given a number value ranging from -3 to 3, though in developing of models, this scoring system was frequently tweaked in order to create a decent model.

Table 7: Incel Trait Statements

Trait	Statement
Defeated	Describes a negative outlook towards any future.
Not confident	Describes shyness and/or low self-esteem.
Rejected	Describes a failure to engage with someone/something.
Confused	Doesn't know what to do, misidentifies.
Sad	Depressed and/or cannot be bothered.
Insecure	Overemphasizing personal faults, weaknesses, and/or failures.
Fearful	Mentions being afraid of someone/something.
Frustrated	Demonstrates feelings of wanting to do something but not being able to.
Excluded	Demonstrates feelings of being outside of some group.
Weak	Describes themselves as physically or mentally inferior.
Scorned	Describes themselves or someone else being mocked.
Shunned	Demonstrates feelings of being intentionally ignored.
Unattractive	Contains descriptions of imperfections in self-presentation, physical appearance, and/or social standing.
Hateful	Contains the use of slurs, hate speech, bigotry, or verbal abuse.
Disgusted	Demonstrates feelings of sickness.
Paranoid	Demonstrates conspiratorial thinking and/or feelings that everyone is out to get them.
Resentful	Demonstrates feelings of being treated unfairly.
Vengeful	Demonstrates motivation to do something about a perceived harm.
Enraged	Is highly aggressive and/or demonstrates potential that they might do something.
Violent	Contains descriptions of physical violence, the desirability of physical violence, and/or fantasies of physical violence.

Alongside the manually scored parent comments, the comments texted were vectorized using Scikit-learn for Python (Pedregosa et al., 2011). Multiple methods of vectorizing were attempted, both using the term frequency-inverse document frequency (TF-IDF) vectorizer used in Topic Modeling, as well as a count vectorizer. Pairing up these vector arrays with the manual scores created, the data was split into a training set and a validation set and the training set was used to train the model once again using Scikit-learn (Pedregosa et al., 2011). Multiple models were tested but none created models with any acceptable validity. As a result, this aspect of the analysis was abandoned.

References

- Allyn, B. (2021). Group-chat app discord says it banned more than 2,000 extremist communities. *NPR*. <https://www.npr.org/2021/04/05/983855753/group-chat-app-discord-says-it-banned-more-than-2-000-extremist-communities>
- Austen, I., & Stack, L. (2018). Toronto van driver kills at least 10 people in ‘pure carnage’. *New York Times*. <https://www.nytimes.com/2018/04/23/world/toronto-van.html>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks [Section: Demonstration Papers]. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1), 361–362. Retrieved June 23, 2022, from <https://ojs.aaai.org/index.php/ICWSM/article/view/13937>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python* (1st ed) [OCLC: ocn301885973]. O’Reilly.
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models [event-place: Pittsburgh, Pennsylvania, USA]. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120. <https://doi.org/10.1145/1143844.1143859>
- Bratich, J., & Banet-Weiser, S. (2019). From pick-up artists to incels: Con(fidence) games, networked misogyny, and the failure of neoliberalism. [Publisher: University of Southern California, USC Annenberg Press]. *International Journal of Communication (19328036)*, 13, 5003–5027. <http://proxy.uchicago.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=139171987&site=eds-live&scope=site>
- Center, N. T. A. (2021). *Hot yoga tallahassee: A case study of misogynistic extremism*. U.S. Secret Service, Department of Homeland Security. Washington, DC.
- Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. *Advanced Visual Interfaces*. <http://vis.stanford.edu/papers/termite>

- Cosma, S., & Gurevich, M. (2020). Securing sex: Embattled masculinity and the pressured pursuit of women’s bodies in men’s online sex advice. *Feminism & Psychology*, 30(1), 42–62. <https://doi.org/10.1177/0959353519857754>
- Ging, D. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities*, 22(4), 638–657. <https://doi.org/10.1177/1097184X17706401>
- Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th python in science conference* (pp. 11–15).
- Hoffman, B., Ware, J., & Shapiro, E. (2020). Assessing the threat of incel violence. *Studies in Conflict & Terrorism*, 43(7), 565–587. <https://doi.org/10.1080/1057610X.2020.1751459>
- Holt, T. J. (2007). Subcultural evolution? examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior*, 28(2), 171–198. <http://proxy.uchicago.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edshol&AN=edshol.hein.journals.devbh28.13&site=eds-live&scope=site>
- Horta Ribeiro, M., Jhaver, S., Zannettou, S., Blackburn, J., Stringhini, G., De Cristofaro, E., & West, R. (2021). Do platform migrations compromise content moderation? evidence from r/the_donald and r/incels. *Proceedings of the ACM on Human-Computer Interaction*, 5, 1–24. <https://doi.org/10.1145/3476057>
- Hu, Y. (2005). Efficient and high quality force-directed graph drawing. *Mathematica Journal*, 10, 37–71.
- Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
- Jaki, S., De Smedt, T., Gwózdź, M., Panchal, R., Rossa, A., & De Pauw, G. (2019). Online hatred of women in the incels.me forum: Linguistic analysis and automatic detection.

- Journal of Language Aggression and Conflict*, 7(2), 240–268. <https://doi.org/10.1075/jlac.00026.jak>
- O'Donnell, K. M. (2021). Incel mass murderers: Masculinity, narrative, and identity. *Ohio Communication Journal*, 59, 64–76. <http://proxy.uchicago.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edo&AN=150697570&site=eds-live&scope=site>
- O'Malley, R. L., Holt, K., & Holt, T. J. (2020). An exploration of the involuntary celibate (incel) subculture online [Publisher: SAGE Publications Inc]. *Journal of Interpersonal Violence*, 0886260520959625. <https://doi.org/10.1177/0886260520959625>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Richardson, L. (2013). *Beautiful soup*. <https://www.crummy.com/software/BeautifulSoup/>
- Scaptura, M. N., & Boyle, K. M. (2020). Masculinity threat, “incel” traits, and violent fantasies among heterosexual men in the united states [Publisher: SAGE Publications]. *Feminist Criminology*, 15(3), 278–298. <https://doi.org/10.1177/1557085119896415>
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In S. v. d. Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 92–96). <https://doi.org/10.25080/Majora-92bf1922-011>
- Van Brunt, B., & Taylor, C. (2021). *Understanding and treating incels: Case studies, guidance, and treatment of violence risk in the involuntary celibate community*. Routledge.
- Waśniewska, M. (2020). The red pill, unicorns and white knights: Cultural symbolism and conceptual metaphor in the slang of online incel communities. In B. Lewandowska-Tomaszczyk (Ed.), *Cultural conceptualizations in language and communication* (pp. 65–82). Springer International Publishing. https://doi.org/10.1007/978-3-030-42734-4_4

- Witt, T. (2020). ‘if i cannot have it, i will do everything i can to destroy it.’ the canonization of elliot rodger: ‘incel’ masculinities, secular sainthood, and justifications of ideological violence. *Social Identities*, 26(5), 675–689. <https://doi.org/10.1080/13504630.2020.1787132>
- Yan, H., Brown, P., & Duke, A. (2014). New details emerge about california killer and his victims. *CNN*. <https://www.cnn.com/2014/05/26/justice/california-killing-spree/index.html>