Exploring Misogyny across the Manosphere in Reddit

Tracie Farrell
tracie.farrel@open.ac.uk
Open University

Jakub Novotny
novotny.jakub@gmail.com
No affiliation

Miriam Fernandez
miriam.fernandez@open.ac.uk
Open University

Harith Alani
h.alani@open.ac.uk
Open University

ABSTRACT

The 'manosphere' has been a recent subject of feminist scholarship on the web. Serious accusations have been levied against it for its role in encouraging misogyny and violent threats towards women online, as well as for potentially radicalising lonely or disenfranchised men. Feminist scholars evidence this through a shift in the language and interests of some men's rights activists on the manosphere, away from traditional subjects of family law or mental health and towards more sexually explicit, violent, racist and homophobic language. In this paper, we study this phenomenon by investigating the flow of extreme language across seven online communities on Reddit, with openly misogynistic members (e.g., Men Going Their Own Way, Involuntarily Celibates), and investigate if and how misogynistic ideas spread within and across these communities. Grounded on feminist critiques of language, we created nine lexicons capturing specific misogynistic rhetoric (Physical Violence, Sexual Violence, Hostility, Patriarchy, Stoicism, Racism, Homophobia, Belittling, and Flipped Narrative) and used these lexicons to explore how language evolves within and across misogynistic groups. This analysis was conducted on 6 million posts, from 300K conversations created between 2011 and December 2018. Our results shows increasing patterns on misogynistic content and users as well as violent attitudes, corroborating existing theories of feminist studies that the amount of misogyny, hostility and violence is steadily increasing in the manosphere.

CCS CONCEPTS

• **Information systems** → **Social networking sites**; *Data mining*; *Web mining*.

KEYWORDS

Hate, misogyny, Reddit, feminist studies

1 INTRODUCTION

The 'manosphere' is a group of loosely incorporated websites and social media communities where men's perspectives, needs, gripes, frustrations and desires are explicitly explored. Women and feminism are typically targets of hostility [14, 32]. In these spaces, discourse tends to revolve around a concept of "men's rights activism", which highlights experiences of discrimination against men, including issues from child custody, to homelessness and forced conscription [14].

The manosphere phenomena has been linked to several prominent, violent crimes perpetrated in the real world by individuals belonging to online communities of self-proclaimed misogynists [1, 2]. These acts were justified, *in the words of the perpetrators*, by a deep hatred for women, whom they perceived as having rejected and betrayed them [32]. These high-profile cases have renewed discussions that surfaced previously during the GamerGate and TheFappening controversies [31], more specifically, about how exposure to misogynistic ideas online may lead to increased violence and threats against women.

Feminist analysis of the manosphere concludes that there is an *ideological shift* away from the men's rights topics that used to unite members toward more misogynistic and violent ideas. Recent discourse analyses of the popular men's rights websites, A Voice for Men[1] and Men Going Their Own Way (MGTOW)[2] point to a backlash toward feminism [31, 32, 38], where even positive sentiments toward rape in some circumstances may be utilised to attract men who feel concerned or excluded by the direction of sexual politics [19].

While these discourse analysis studies provide in-depth observations of misogynistic rhetoric, they are usually conducted over a small subset of conversations, and over a very limited period of time. Full manual analysis is impractical and thus, automatic techniques need to be used.

Existing automatic methods to analyse online misogyny are, however, scarce and mainly focused on the analysis of Twitter (a platform where communities are not well-defined) and conducted by using a snapshot of tweets (generally in the thousands) collected over a few months. In addition, except for the work of Hardaker and

---

[1]https://www.avoiceformen.com/
[2]https://www.mgtow.com/

McClashan [20], existing computational studies do not usually build over the extensive knowledge and years of research from discourse studies and feminist linguistic models of online misogyny.

In this paper, we triangulate our investigation of misogyny online between feminist theories, social theories, and computational methodologies to gain a better understanding of misogyny online. In particular, we explore how extreme and violent language, specifically relative to women, is expressed and how it evolves across seven different communities from Reddit[3], where users group around shared interests, ideologies, and subcultural language. Our study is conducted over 7 online communities including 6M posts grouped around 300K conversations, which, to the best of our knowledge, forms the largest study of this topic so far. Our research is guided by the following research questions:

- *What is the strength and evolution of misogynistic ideas within and across different online communities?* To tackle this question we propose an approach based on Natural Language Processing that captures 9 prevalent misogynistic ideas (or categories of misogyny) from feminist studies and translates these categories into lexicons to automatically track their emergence and evolution within and across communities.
- *Which groups express the most violent attitudes and what are the most popular forms of online misogynistic violence?* To conduct this study we assess communities considering the amount of content and users expressing violent attitudes, the type of violent attitudes expressed, and the key terminology used for this purpose.

By investigating these research questions, we provide the following contributions:

- A summary of a wide range of theories and models of online misogyny from the feminist literature, as well as an analysis of the works that have targeted the problem of online misogyny from a computational perspective.
- The translation of different categories of misogyny, identified in feminist theory, into lexicons of hate terms to study the evolution of language within the manosphere.
- An in-depth analysis of different manifestations and evolution of misogyny across the Reddit manosphere.
- We corroborated existing feminist theories and models around the phenomenon of the manosphere by conducting a large-scale observational analysis.

The rest of the paper is structured as follows: Section 2 describes related work, including feminist models of online misogyny, linguistic models of online misogyny and computational approaches. Section 3 shows our proposed approach to operationalise well known manifestations of misogyny from social science and feminist studies to lexicons of terms and expressions that we can track computationally. Sections 4 and 5 describe our analysis of the manosphere on Reddit.Section 6 discusses our results and implications, and Section 7 concludes the work.

2 RELATED WORK

Misogyny is hatred or contempt for women [8]. Yoon described misogyny as "the police force of sexism"[37], linguistically and

---

[3]reddit.com

behaviourally subjugating or excluding women within what appears to be a patriarchal society [8]. Misogyny is often positioned in opposition to feminism, and as evidence for the ongoing need for feminism to continue [5, 14]. From a feminist perspective, misogyny is *visibly carried out online*, evidenced by the documented hate that has been targeted specifically at women celebrities, politicians, and other professionals for no apparent reason other than their gender. To provide two high-profile examples, nude photos of celebrities leaked during what was called "The Fappening" affected almost exclusively women. It was also largely women in the gaming industry, who were targeted with violent messages and threats during the #GamerGate controversy, in response to what was perceived by some men as undue progressiveness in gaming [31].

Using violent and misogynistic language online is not trivial, even if some individuals do not intend to act on any of their statements. Online men's groups have given space to members' glorification of tragedies like the Isla Vista Killings, in which Elliot Roger killed 6 people and wounded 14 others, after communicating extensively online about his contempt for women (and people of colour) [1]. That rhetoric is believed to have attracted others, including Alek Minassian, the alleged perpetrator of the Toronto van attack that killed 10 and wounded 16 people [2]. Along with the #GamerGate and #TheFappening controversies, which impacted hundreds of women [31], there are growing concerns that misogyny *online* has some worrying qualities in scope and scale that women are unable to avoid [25].

In this section, we discuss some of the ways in which feminist scholars have attempted to shed light on the issue of misogyny online and the limitations of these approaches with regard to scope and scale. We also explore existing computational approaches and present some of the challenges that those approaches have not yet addressed.

2.1 Feminist Models of Online Misogyny

Feminist scholarship explores misogyny online through research questions that acknowledge culturally and socially embedded experience [7, 8, 23], including:

- the nature of misogyny and attempts to characterise it
- exacerbating factors, both social and technological
- impacts on society and culture

Feminist theory explores misogyny historically [8], as well as in contemporary contexts, to demonstrate how misogyny evolves alongside culture [14]. Misogyny is presented not only as behaviour that objectifies, reduces, or degrades women, but also as the exclusion of women, manifesting itself in discrimination, physical and sexual violence, as well as hostile attitudes toward women [8]. Feminist scholarship argues that framing misogyny in this way is important for examining the overarching structures and conditions that allow misogyny to persist [29, 37]. Defensive hashtags, such as notallmen or FemalePrivilege, imply that because not all men are misogynist and some women are more powerful than some men, misogyny does not really exist as part of cultural or political hegemony [18].

However, feminist studies on the nature of misogyny online are typically conducted by a small number of authors looking at a small amount of data intensively. For example, both Kendall [26] and Lin

[28] joined online groups to observe and interview group members about misogynistic views (Kendall joined what she called the BlueSky community, Lin joined MGTOW). Lin builds her argument through comparing and contrasting community members' statements with one another, and contextualising them within a timeline of current events [28]. Kendall spent 3 years as a member of the group, having informal contact as well as conducting interviews with other members [26]. The data from such studies is very rich, but it is insufficient to *structure* deeply entrenched or networked misogyny, as we see online [5].

Zuckerberg investigated networked misogyny through her analysis of rhetoric on the 'manosphere' and the ways in which misogyny is justified within it. She found that these groups often shared an exploitation of "deeply misogynistic" ancient, Hellenistic philosophy that allow them to reframe history and promote violence against women. According to Zuckerberg, the manosphere flips the narrative of oppression and positions men both as victims of exclusion under the authority of feminism [38]. However, Zuckerberg's analysis is based on an inductive logic of community discourse, making it difficult to understand how deeply she examined communities and for how long.

Misogyny is exacerbated by both social and technological factors, according to feminist scholars [30]. Feminists connect the disenfranchisement of women with many other categories of exclusion, such as race, ethnicity and sexual orientation [8, 10]. Anonymity and technology affordances are believed to play an further aggravating role in this phenomenon, by giving perpetrators a way of voicing extreme sentiments without taking responsibility for their words or their consequences [5]. Stoeffel, for example, analysed case studies involving women's experience of violent threats online. She argued that anonymity not only disproportionately impacts women, it puts professional women in particular danger, as they engage as their real selves against anonymous abusers [35]. From a cyberfeminist perspective, these events signify that "toxic technocultures" have developed online and their power is being used to exclude, humiliate, extort and injure women [31].

Feminist studies about misogyny online measure impact by looking at the types and volume of hate abuse received [35], spikes in abuse accompanying perceived "gains" in feminism [5], but also the impact on men, who also experience hostility in the form of feminisation of their own character or insults to female members of their family [24]. Methods of feminist research in this area are typically observational or case studies, tracing women's behaviours and sentiments across various online platforms, and capturing the violent responses to this from men [24]. These studies are discursive and illuminate important relationships and dependencies. However, for a high level picture of the evolution of such relationships over time, computational approaches could provide valuable evidence.

Massanari, for example, studied the infrastructure of the platform Reddit, the same platform we studied in our research. She found that Reddit's karma point and subreddit systems, ease of account creation and loose governance structure/policies were creating an environment for "toxic technocultures" to proliferate [31]. While this study included a manual historical analysis of activity across the subreddits she was observing, Massanari's research is limited

without a way of systematically observing the evolution of communities over a clear period of time, which could help interrogate her theories of how toxic masculinity spreads online.

2.2 Linguistic Models of Online Misogyny

Feminist theorists of language and gender address many themes, from different uses of language by men and women, socialisation through language, the power of definition of terms, as well as language and identity construction [11]. The assumption at the heart of each of these themes, is that power determines who sets the rules for language and that language can therefore be used to control or subjugate. Feminist studies on the use of language on the internet should be viewed through this lens - that language is political as it is personal and cultural.

Sundén and Paasonen, for example, examined different pejorative terms for women of a certain age and ideology, which women have now attempted to "reclaim" in the Swedish and Finnish context through discourse analysis [36]. Marwick and Caplan performed a critical discourse analysis on the word "misandry" (hatred or contempt for men) as it evolved through different online communication platforms from the 1990s to 2010s. They searched archives within Google Groups to search for misandry by year within different communities, as well as Google Trends, the Internet Archives' WayBack Machine, and MediaCloud, looking for spikes in usage, which they then examined in more granularity. They applied feminist discourse theory and grounded theory in the analysis of their data, to see how the concept has infiltrated the mainstream and the impact on perceptions of feminism as "a manhating movement which victimises men and boys" [30].

These two linguistic examples demonstrate a fruitful area where computational techniques can improve upon and enhance existing approaches, providing more efficient ways of identifying some types of anomalies and providing a historical picture of the evolution of language and activity over time.

2.3 Computational Approaches

While the problem of online hate speech has been the focus of a wide body of research during the last few years [15], computational approaches targeting the problem of misogyny in particular are scarce and very recent. Computational methods have been either used to **observe** and study the phenomenon of online misogyny [6, 20–22], to generate automatic misogynistic **content detection** methods [4, 12, 13], or to use the appearance of misogyny related words in online content as a **predictor** of criminal behaviour [16].

Regarding the **observational studies**, Jamie Bartlett and colleagues [6] analysed misogyny on Twitter by collecting English tweets containing the word 'rape' during a three month period between December 2013 and February 2014. Based on a sample of 138,662 tweets they studied the over time evolution of this word, and differentiated between casual and offensive misogyny.

Hardaker and McClashan [20] studied the case of online misogyny against the feminist campaigner Caroline Criado-Perez. 76,275 tweets were collected during a three month period (25/06/13 - 25/09/13) and quantitative and qualitative approaches were used to detect emergent discourse communities. As the authors pointed out,

this notion of community is based on users who have no connection to each other rather than either supporting or abusing Caroline.

Hewitt and colleagues [21] gathered 5,500 tweets over the course of a week based on a set of 20 terms and annotated those tweets manually (each tweets was coded by one researcher as misogynistic or not). They used this exercise to discuss the problems of identifying misogynistic language on Twitter and other online social spaces, since it is not always trivial to determine what is misogynistic language. They propose the use of clustering analysis as a way to automatically mine the language that emerges from the data.

Jaki and colleages [22] conducted a study of the online discussion forum Incels.me and its users. They analysed 65,000 messages posted during a 6-month period between November 2017 and May 2018. They identified the terms more frequently used by the community and did a qualitative analysis to identify key topics of discussion. Their analysis highlights that about 30% of the content is misogynistic, 15% homophobic and 3% racist.

One step further from observational studies, Maria Anzovino and colleagues [4] focus on the **automatic detection and categorisation** of misogynous language in social media. They design a taxonomy of manifestations of misogyny that includes five different categories: discredit, stereotype and objectification, sexual harassment and threats of violence, dominance, and derailing. They collected tweets based on the set of words proposed by Hewitt and colleagues [21] from July till November 2017. 2,227 tweets were annotated as (misogynistic or not) and (as belonging to one particular category). This dataset was further used to generate automatic identification methods. Extensions of this dataset, including English, Spanish and Italian tweets have been built by the authors and used within two competitions where the challenge is to create classifiers to automatically identify misogyny online; IberEval-2018 and EvalIta-2018.[4] This has lead to a series of papers investigating various classification methods to categorise misogyny including: Support Vector Machines (SVM), Logistic Regression, Ensemble Models, or Deep Learning [12, 13].

In addition to these studies Fulper and colleagues [16] investigated the potential of social media in **predicting** criminal behaviour, in particular rape and sexual abuse. For this purpose they compared the volume of misogynistic tweets and the rape crime statistics in the United States, finding a significant association. To identify misogynistic tweets they compiled a list of 90 terms (not described in the paper and, to the best of our knowledge, not publicly available). These terms were used to categorise 1.2Million tweets as containing misogynistic language.

As we can observe, our proposed study differentiates from previous works in several important directions. First of all, while our approach is also observational, our aim is not to study a particular use case [20], or a small sub sample of tweets collected during few months [6, 21, 22], but concrete communities, where users share in-group characteristics, like common ideology or subcultural language. Moreover, we do not aim to observe a time snapshot, but the full evolution of these communities from their inception. Our study is conducted over 7 online communities including 6M posts grouped around 300K conversations, which, to the best of our knowledge, is a significantly larger study than previous ones. The platform used

for our study is also different than the one used by previous works (Twitter). Not only Reddit enables the creation of communities around shared interested, it also allows for longer posts (not just the 280 characters of Twitter), providing a richer ground to analyse and explore linguistic phenomena. It is also important to highlight that, except the work of Hardaker and McClashan [20], existing computational works do not take advantage of the knowledge and years of experience from discourse studies and feminist linguistic models of online misogyny when conducting their analyses or proposing detection and prediction methods.

3 REPRESENTING MISOGYNY

Feminist theory describes misogyny as a range of activities from hostility toward women, to physical, psychological and systemic violence against them. The Encylopedia of Feminist Theories [8] identifies six key activities that are connected with misogyny: (i) Physical violence towards women, (ii) Sexual violence towards women, (iii) Hostility towards women, (iv) Belittling of women (v) Exclusion of women and (vi) Promotion of Patriarchy or Male Privilege. Adding to these basic characteristics, we add those that arise in connection with misogyny online: (vii) Stoicism (from Zuckerberg) and, (viii) Flipping the Narrative (from Ging [18] and Flood [14]).

In this work, we aimed at linguistically characterising these manifestations of misogyny by building lexicons of hate with the terms and expressions that describe these particular ideas or categories of misogyny. We built a total of 9 lexicons including the above mentioned categories and two additional ones that encapsulate other hateful terms and expressions not necessarily targeted towards women, but that aim to capture the distinct levels of hateful speech within the manosphere. These categories are Homophobia and Racism. To build such lexicons we used seven existing lexicons of hate speech and studies of the specific rhetoric used within the manosphere. These list of lexicons include:

- *Harassment Corpus*. Built by Rezvan and colleagues[33],[5] this lexicon includes an annotated corpus of 713 harassment words and expressions including: sexual harassment (452 terms), racial harassment (153 terms), appearance-related harassment (14 terms), intellectual harassment (31 terms), political harassment (21 terms) and general (42 terms)
- *Violence verbs*: Built by Geen and Sonner [17], this lexicon encapsulates 322 verbs identified with several categories of violence including: torture, stabbing, murder, massacre and choke.
- *Hatebase (female)*: Hatebase is the world's largest structured repository of regionalised, multilingual hatespeech.[6] It contains 2,432 terms in 94 languages associated with hate. From this list, using Hatebase search capabilities, we filtered 36 hate terms commonly used towards females.

Table 1: Lexicons of Misogyny

Category of Misogyny	Num terms	Examples
Belittling	58	femoid, titties, stupid cow
Flipping the narrative	7	beta, normie, men's rights
Homophobia	126	dyke, fistfucker, faggot
Hostility	303	bitch, cunt, whore
Patriarchy	8	alpha male, subjugate, suppress
Physical Violence	73	hit, punch, choke
Racism	670	nigger, raghead, pikey
Sexual Violence	22	rape, sodomise, gangbang
Stoicism	33	blackpill, cuck, hypergamy

- *Hatebase (original)*: Davidson and colleagues [9] published with their work 1,034 terms from the original Hatebase dictionary.[7] To the best of our knowledge the current full Hatebase dictionary is not publicly available.
- *Profanity words*: Published by Robert J. Gabriel, this lexicon contains a list of 450 bad words and swear words banned by Google.[8]
- *Incel specific*: This dictionary, created by Tim Squirrel, contains an analysis of 57 specific neologisms from the Incel (Involuntary Celibates) community.

The collected 2,454 unique terms and expressions have been manually coded by the first author of this paper (native English speaker with a background in social science and cultural anthropology). Out of this annotation 1,300 terms have been selected, since they belong to one the above mentioned categories of misogyny. The created lexicons are available here.[9] The number of terms per category is listed in Table 1. To annotate these terms the following considerations were taken into account:

- The concept of stoicism is based on Zuckerberg's analysis of the manosphere [38]. It encapsulates terms and expressions of endurance of pain or hardship because of the lack of intimacy or beauty. Terms like 'kiss-less', 'hug-less' or 'involuntarily celibate' are part of this category.
- Patriarchy encapsulates that which has to do with women being considered less than men, or some men being better than others by virtue of having traditionally masculine qualities.
- Flipping the Narrative encapsulate terms and expressions that refer to men being oppressed by women or (indirectly or directly) by other men.
- Sexual Violence encapsulates any word explicitly connected with sexual violence (and nothing else).
- Physical Violence encapsulates any word explicitly connected with physical violence that is not explicitly sexual.
- Hostility includes violent verbs, and slurs that are not immediately racist or homophobic. However, if a verb is ambiguous (such as fucking), but it is made into a slur (such as fucker) it is coded it as hostility.
- Belittling encapsulates any word that is disrespectful or degrading women's experiences.

---

[7]https://github.com/t-davidson/hate-speech-and-offensive-language/tree/master/lexicons
[8]https://github.com/RobertJGabriel/google-profanity-words-node-module/blob/master/lib/profanity.js
[9]https://github.com/miriamfs/WebSci2019

- Homophobia encapsulates any word related to being homosexual or that mocks being homosexual. This category does not distinguish between terms that have to do explicitly with women and those that have to do with men, for reasons that this was a modifying category to assess general violent attitudes.
- Racism referred to any word that was supposed to represent a specific group of people based on where they are from or their perceived race or ethnicity. Xenophobia and racism are not separated in this category. We included terms like 'fresh off the boat', 'paddy' (pejorative for Irish person) and 'kraut' (pejorative for a German person), as well as terms referring specifically to race. The reason for this division is, similarly to the above, to simplify modifying categories about violent attitudes.

4 ANALYSIS SET-UP

In this section we describe the different online communities selected for this study as well as the analysis conducted to answer our research questions.

4.1 Selection of Online Communities

To choose our communities, we began with one community that had been in the media following the Toronto van attack, Incels [3]. Critical discourse analyses of these groups' communications online had uncovered predator-prey or master-servant dichotomies in their discussions about male-female relationships, and a reinforced sense of entitlement [3, 27]. For Incels, misogyny appeared to be coupled with racism as well, in particular for women of colour who are perceived as racially betraying darker skinned men in favour of white men [3, 34].

From this node, we intended to find a selection of related communities that have some shared ideologies. In April 2018, we identified 10 groups with the word "incel" in their community name or in their discussion threads, which were still active and online on Reddit. Of those 10 groups, we saw that 1 group was private, 4 groups were monitoring groups and 5 groups were active, self-identified groups of incels or men's rights activists. To help understand how such groups might differ from one another, we collected data from many different types of communities discussing some of the same information and ideas. In particular, to conduct our analysis we collected information from six popular communities on Reddit, which revolve around topics expressed in the manosphere, such as men's rights and difficulty with relationships:

- *r/MGTOW:* this is a subreddit of 'men going their own way', in which men claim that they wish to simply live a life without the interference from women.
- *r/badwomensanatomy*: this is a subreddit focusing on women's bodies in a misogynistic way.
- *r/Braincels*: this is the main incel subreddit since *r/incels* was removed from Reddit in November 2017 for violating site-wide rules. It is widely believed that this happened because of a post from an *r/incels* user about legal advice in which he pretended to be asking a "general question about how rapists get caught". Some members of Braincels were also

self-reported members of the website incels.me (now defunct), and the more current incels.is[10] or similar non-Reddit websites, where more violent content is posted.

- *r/IncelsWithoutHate*: this is a subreddit of individuals who are self-described as both incel but non-violent. This group supplies[?] somewhat of a control group, in that they will share some of the same vocabulary with other groups, but should express less misogyny and violence as other groups.
- *r/Inceltears*: this is a subreddit dedicated to calling out Incels. They screenshot and post particularly egregious content from r/braincels, incels.me, incels.is and other incel communities. They are partly responsible for a large number of incel communities being closed down.
- *r/IncelsInAction*: this is a subreddit that monitors activity from other incel communities, similarly to *r/Inceltears*.
- *r/Trufemcels*: this is a subreddit of women who are self-described incels. Male incels occasionally remark that it is not possible for a female incel to exist, given the advantages of women over men in finding a sexual partner.

4.1.1 Data Collection. We gathered all data from the above seven communities, from their inception until January 2019. This has led to the collection of 301,078 conversations and a total of 5,674,303 comments in those conversations (see Table 2). We collected this data via the pushshift API[11], a big-data storage project that maintains an archive of Reddit data. Table 2 shows a summary of the collected data, including the online community, its number of posts, and the dates of the first and last post. We collected a data snapshot from the beginning of each community until 11/01/2019. Note that for Braincels we have data until 01/10/2018, since this community has been put in quarantine by Reddit. We can observe that MGTOW is the largest community, in terms of contributions (nearly 200K posts) and also the one that has been active for longer (since June 2011). The most active community is however Braincels, which contains nearly 100K contributions done in one year from October 2017 till October 2018. The less active community is IncelsInAction, with only 330 posts, and the newest one is Trumfemcels, which originated in April 2018.

4.2 Conducted Analyses

We used the constructed lexicons to identify the amount and type of misogynistic content posted in each community. Let's C be the set of communities, P be the set of posts, and P_{c_i} the set of post

---

[10]https://incels.is/
[11]https://pushshift.io/

Community	numPosts	minDate	MaxDate
MGTOW	168124	2011-06-04	2019-01-11
badwomensanatomy	13010	2014-01-02	2019-01-11
IncelsWithoutHate	2309	2017-04-09	2019-01-11
IncelTears	15679	2017-05-19	2019-01-11
IncelsInAction	330	2017-06-24	2019-01-10
Braincels	96545	2017-10-21	2018-10-01
Trufemcels	5081	2018-04-04	2019-01-11

Table 2: Table captions should be placed below the table.

of a particular community c_i. Let's M be the set of misogynistic categories considered in this study, L the set of lexicons and l_j the lexicon that represents the category of misogyny m_j. We consider that a post $p \in P_{c_i}$ can be labelled as displaying the category of misogyny m_j if it exists a term t in the post p that belongs to l_j, i.e., $t \in p$ and $t \in l_j$. It is also important to notice that for the purpose of this study we do not differentiate between initial posts (i.e., those that start a new thread) and comments (i.e., contributions to an existing thread). Based on this proposed mapping of content to categories of misogyny we have conducted several analyses to answer our research questions:

- Table 3 shows the amount of misogynistic posts for each community and the distribution of these posts across categories. Percentages are provided as well as totals to assess the relative strength of each misogynistic category within each community.
- Table 4 shows the amount of users posting misogynistic content for each community and for each category within each community. Percentages are provided as well as totals to assess the relative size of the members of the community engaged in misogynistic behaviour.
- Table 5 shows the top terms used across all communities to describe the different categories of misogyny.
- Figure 1 displays the over time evolution of the different categories of misogyny for every community.
- Figure 2 displays the over time evolution of the number of active users for every community.

5 ANALYSIS RESULTS

In this section we display the results of our analyses and discuss key insights.

5.1 Strength and Evolution of Misogyny

Following RQ1 (see 1) this analysis investigates the strength and evolution of misogynistic ideas within communities. Table 3 shows a static snapshot of the number of posts for each community, the number of posts displaying misogyny (based on our lexicon-based approach), and the number of posts displaying each specific category of misogyny. Totals and percentages are provided in this table. In addition, Figure 1 displays the evolution of each category of misogyny within each of the online communities analysed.

As we can see from Table 3, while MGTOW and Braincels display the higher number of posts categorised as misogyny (726K and 451K respectively, since they are the biggest communities, in terms of percentages of misogynistic conversations, Braincels, IncelsInAction, IncelTears and IncelsWithoutHate all display more than 27%. In terms of specific categories of misogyny TruFemcels shows the highest percentage for belittling and racism, MGTOW shows the highest percentage for flipping the narrative, hostility and physical violence, Braincels shows the highest percentage for homophobia and patriarchy, IncelTears for sexual violence and IncelsWithoutHate for stoicism. We can also observe that Hostility, Stoicism and Physical violence, seem to be the most popular categories across communities.

These results can be triangulated with how these communities self-identify to further examine what we have observed. For example, members of IncelTears and IncelsInAction are reporting on extreme posts that they see on other subreddits where incels are active (such as Braincels). We would expect to see a higher concentration of misogynistic language on that subreddit, which we do. Similarly, however, we would expect IncelsWithoutHate, with the community's aim to spread *less* violent speech than other groups of incels, to have a lower percentage of some forms of misogyny than other groups. However, IncelsWithoutHate had the greatest percentage of posts including misogynistic language. Returning to the actual content on the subreddit, we observed that IncelsWithoutHate reference content they have seen posted by other more violent groups (similarly to IncelTears and IncelsInAction). There also appears to have been some infiltration of the group and harassment[12].

TruFemcels, as a community for women incels, is also difficult to interpret. Misogyny is not exclusive to men. Feminist scholars view internalised misogyny as evidence of patriarchy [6]. However, the explanation for this may also have to do with interference on the subreddit from agitators. Moderators of the subreddit have warned that TruFemcels are regularly harassed by what is believed to be male incels. This harassment include doxxing[13] and trolling threads.[14]

It is important to note that this is a static view of the communities (considering their full duration). We also analysed their evolution over time to observe the dynamics of these misogynistic ideas. The over-time (weekly) distributions are displayed in Figure 1. The first interesting element we observed is a significant difference in the time/posts distributions among the different communities. For MGTOW, categories of misogyny seem to have a similar evolution pattern, being very mild or nearly nonexistent during the first two years of the community, but displaying a constant increase since 2015 till the end of 2018 were we observe a slight decrease. This growth is parallel to the growth on the number of active members posting in the community (see Figure 2). Within this community hostility is the most prominent category, followed by physical violence and belittling. This is a similar pattern to that displayed by the subreddit badwomensanatomy with several exceptions. Categories are more interconnected (i.e., they are less prominent with respect to one another) and activity in the community (number of posts) is smaller. However, hostility, physical violence and belittling are also the most prominent categories.

For Braincels, we can observe a steady increase for all categories of misogyny, and a sharp increase around April 2018. One possible explanation for this may have to do with the Toronto van attack, allegedly perpetrated by self-proclaimed incel Alek Minassian during this same period [2]. We see peaks in other communities as well at this time. From our investigations, this could be extended media attention on the subject of incels, following the Toronto attack. Searches for the term incel on google trends shows spikes from April and November 2018 as well. After that, the weekly amount of misogynistic content produced on the subreddits we examined

stabilises. The top categories of misogyny that particularly stand out in this community are hostility and stoicism.

IncelsInAction, IncelsWithoutHate, IncelTears and TruFemcels display more constant patterns of growth with hostility and stoicism being prominent categories in all of them. For the first three groups, this is expected, as they are involved in re-posting content they view as particularly hateful or demeaning. For TruFemCels, this is once again difficult to interpret, considering the harassment that the group has received. However, after returning to the community to examine some usage of words in more detail, the high levels of stoicism appear to accurately characterise this group.

MGTOW, Braincels, IncelsWithoutHate and TruFemCelsFrom are the communities we studied that are actually for the incel or men's rights communities. In our analysis, we see stoicism in the four groups we examined increasing over the course of 2017 and 2018, with a slight downturn in late 2018. This supports Zuckerberg's proposition that stoicism is an important, over-arching narrative in these communities [38]. Interestingly, we also identified this characteristic in TruFemCels, where misogynistic attitudes are more related to belittling and self-deprecation.

It is also interesting to observe the steady and sometimes sharp growth of active members in these communities (see Figure 2). The activity within MGTOW and Braincels have been steadily growing, reaching picks of nearly 5,000 active users a week. For Braincels we can perceive a major and rapid increase of active members at the end of April 2018 followed by a sharp decrease around the end of the year, when the community was quarantined by Reddit. Badwomensanatomy and inceltears also display a steady increase, with an actual activity status of 3,000 active users every week. IncelsWithoutHate, Trufemcels and IncelsInAction display more moderated patterns of growth.

These findings, and our observations of the constant increase in the different categories of misogyny across all communities, support existing discourse analysis studies that violence and hostility are increasing toward women online [5, 30, 31].

5.2 Violent attitudes

Tables 3 and 4 display the amount of misogynistic content and the number of users spreading misogynistic content for every community. As we can see in this analysis, MGTOW, Braincels and badwomensanatomy (all groups that are actual incel or men's rights communities), are the communities with higher amounts of misogynistic content and users. Massanari's idea of "toxic technocultures" [31] can possibly be seen in this data, as the number of users posting misogynistic content and the amount of misogynistic content are both increasing across our groups of male incels and men's rights activists.

In terms of violent attitudes, (which we identify with the categories of sexual violence, physical violence, racism and homophobia) we can observe that for sexual violence, MGTOW and Braincels display the largest amount of posts, but IncelTears and IncelsInAction display the highest percentages. Again, as monitoring groups, we would expect to see a high level of misogynistic content in these subreddits, which has possibly come from MGTOW or Braincels. In terms of physical violence, MGTOW displays both the largest amount and the highest percentage of posts, followed by IncelTears

---

[12]https://tinyurl.com/y2cgbfav
[13]finding another users private information and making it public online
[14]https://tinyurl.com/y6z234r5

Community	N.Posts	Misogyny		Belittling		FlippingNarr		Homophobia		Hostility		Patriarchy		P. Violence		Racism		S. Violence		Stoicism	
		N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P
MGTOW	3136231	726621	23.17%	97039	3.09%	52755	1.68%	3399	0.11%	323867	10.33%	1799	0.06%	125916	4.01%	38444	1.23%	38408	1.22%	44994	1.43%
badwomensanatomy	496665	72046	14.51%	18278	3.68%	367	0.07%	237	0.05%	28003	5.64%	132	0.03%	12495	2.52%	7215	1.45%	3101	0.62%	2218	0.45%
IncelsWithoutHate	54789	16307	29.76%	2082	3.80%	626	1.14%	85	0.16%	4718	8.61%	51	0.09%	2114	3.86%	847	1.55%	455	0.83%	5329	9.73%
IncelTears	578659	161732	27.95%	17003	2.94%	3505	0.61%	903	0.16%	54404	9.40%	355	0.06%	22293	3.85%	9415	1.63%	12263	2.12%	41591	7.19%
IncelsInAction	3903	1055	27.03%	102	2.61%	8	0.20%	7	0.18%	340	8.71%	1	0.03%	121	3.10%	51	1.31%	75	1.92%	350	8.97%
Braincels	1629277	451754	27.73%	47489	2.91%	21384	1.31%	9218	0.57%	149179	9.16%	1279	0.08%	37766	2.32%	31944	1.96%	10048	0.62%	143447	8.80%
TruFemcels	75553	18045	23.88%	2940	3.89%	588	0.78%	72	0.10%	5608	7.42%	60	0.08%	2326	3.08%	1892	2.50%	326	0.43%	4233	5.60%

Table 3: Misogynistic content across categories and communities. N=Number of posts. P=Percentage of posts

Community	N.Users	Misogyny		Belittling		FlippingNarr		Homophobia		Hostility		Patriarchy		P. Violence		Racism		S. Violence		Stoicism	
		N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P
MGTOW	93596	33879	36.2%	14339	15.32%	9627	10.29%	1606	1.72%	25541	27.29%	1112	1.19%	16522	17.65%	9243	9.88%	8273	8.84%	9507	10.16%
badwomensanatomy	69076	19798	28.66%	8702	12.6%	310	0.45%	208	0.3%	11616	16.82%	115	0.17%	6433	9.31%	4379	6.34%	1982	2.87%	1536	2.22%
IncelsWithoutHate	5198	2205	42.42%	786	15.12%	330	6.35%	71	1.37%	1334	25.66%	40	0.77%	829	15.95%	439	8.45%	248	4.77%	1316	25.32%
IncelTears	56273	22867	40.64%	6725	11.95%	2058	3.66%	674	1.2%	13952	24.79%	284	0.5%	7861	13.97%	4340	7.71%	4821	8.57%	11615	20.64%
IncelsInAction	1642	496	30.21%	94	5.72%	9	0.55%	7	0.43%	236	14.37%	1	0.06%	98	5.97%	46	2.8%	57	3.47%	237	14.43%
Braincels	55516	22968	41.37%	8658	15.6%	5042	9.08%	2620	4.72%	16089	28.98%	681	1.23%	8599	15.49%	6652	11.98%	3554	6.4%	13899	25.04%
TruFemcels	7727	2680	34.68%	1017	13.16%	294	3.8%	59	0.76%	1506	19.49%	48	0.62%	879	11.38%	688	8.9%	174	2.25%	1305	16.89%

Table 4: Misogynistic users across categories and communities. N=Number of users. P=Percentage of users

Belittling		FlippingNarr		Homophobia		Hostility		Patriarchy		P. Violence		Racism		S. Violence		Stoicism	
word	freq	word	freq	word	freq	word	freq	word	freq	word	freq	word	freq	word	freq	word	freq
female	121549	mgtow	38862	faggot	7453	hate	89389	suppress	1188	hit	37142	black	49552	rape	58396	incel	111373
dumb	31312	beta	33450	fag	1856	bitch	54609	betabuxx	1023	kill	29006	bigger	15545	cock carousel	3750	chad	36081
roastie	9143	normie	16975	fags	1407	pussy	47869	omega	786	cut	21888	nigga	3575	pound	2703	cuck	31482
boobs	8354	mra	424	homo	1278	hurt	26440	subjugate	300	force	17740	nigger	1030	conquer	1543	cope	26169
failure	7952	overthrow	368	dyke	879	cunt	24150	oblige	255	attack	12061	niggas	1006	incest	1378	blackpill	13621

Table 5: Most used misogynistic terms across communities

(a monitoring group). Braincels is largely the most homophobic community and TruFemcels the one that displays a higher percentage of racist content. In terms of users, MGTOW shows the highest number of users talking about sexual and physical violence, while Braincels shows the highest percentage of users displaying homophobic and racial language.

Among the most used violent terms across communities 5 we can observe hit, kill, cut, and rape occurring with a high level of frequency in comparison to other words. One exception to this are neologisms that developed in the communities. The words incel, chad (normatively attractive white male) and blackpill (deterministic point of view on human relationships) are examples. Another exception is the collection of words that the communities have appropriated. Beta, for example, is a term for a man who is not an alpha male (white, attractive and successful). The word cuck has many different meanings from our examination of the subreddits, but broadly refers to anyone who does not accept what is perceived to be biologically determined hypergamy among women. The development of new language is another signal of the existence of a "culture" of misogyny that may exist within these online communities in the manosphere [5, 31].

6 DISCUSSION

With this observational analysis our goal has been to test existing feminist theories and models at scale. Our results do indeed corroborate some of these theories, particularly, the idea that violence and hostility are increasing towards women online [30], that violent rhetoric and misogyny are co-occurring [8, 10], and that stoicism and flipping the narrative are two contemporary responses to feminism [38]. Our research supports these positions by exploring the evolution of content and users over time in seven different communities. While we cannot indicate a clear motive for violence and hostility, we can say that it is increasing, that fluctuations exist in the misogynistic language used by these groups, and that stoicism and hostility are increasing and prominent across communities.

One limitation of using lexicons in this study is that they are unable to capture all of the words that might be relevant (lack of completeness). They also do not capture important details about the context of language [6, 21]. Despite these limitations, these lexicons helped in exposing the trends over time in the use of language. Indeed, we were able to support and extend the hypothesis of feminist studies that the amount of misogyny, hostility and violence is steadily increasing in the manosphere.

We utilised the efforts of only one coder to annotate the words in the lexicons. This could lead to potential bias or human error in annotating the data, and is subject to further work. Nevertheless, we
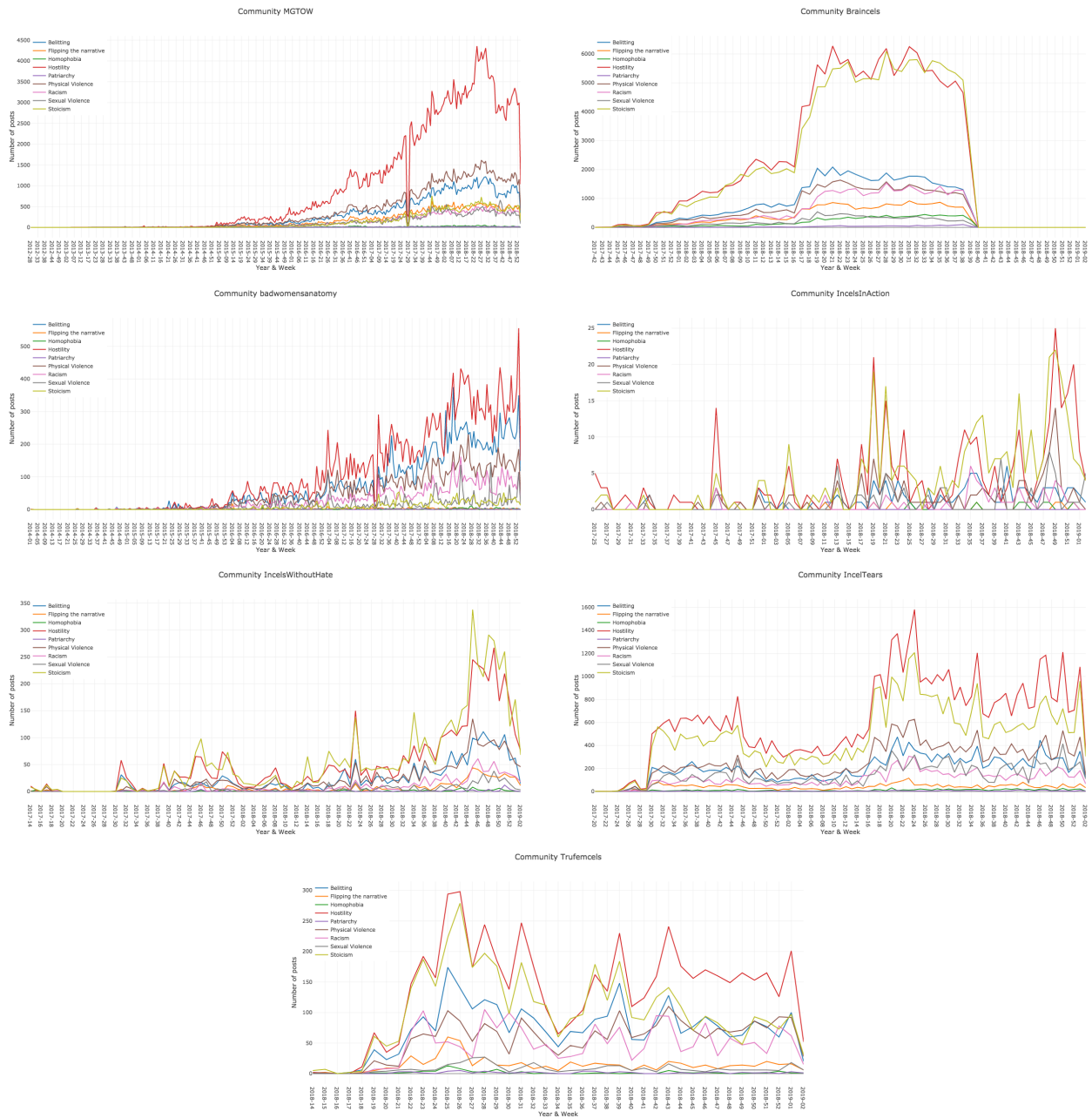
Figure 1: Distribution of misogynistic content posted per week for every category and community

believed it was more critical to have a person familiar with contemporary feminist theory working with the data and discussing codes with the rest of the research team. In the process of our analysis, we identified ways of optimising our lexicons by combining or adjusting categories, some of which we have already implemented and will continue to improve to make these lexicons an open resource for the research community.

In some categories, such as patriarchy or flipping the narrative, there were very few words available in the lexicons. It is possible

that some categories may be disadvantaged towards others, but we can see trends over time despite this limitation. Patriarchy, in particular, was also a difficult category to identify. It does not, therefore, feature prominently in our analysis at this time.

Neologisms may also play an important role, as we found a lexicon of many new terms originating in the incel and men's rights communities. Although 57 studied neologisms were incorporated in these lexicons, through the study of these communities we have detected newly emerged neologisms that we plan to study in the
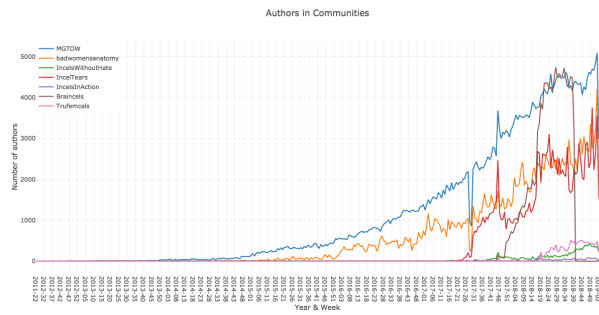
Figure 2: Weekly active users across communities

future. In particular, we are interested in where these new terms emerge, how they are used by different communities and where divergence and similarities can be observed. This forms part of our future work.

Finally, we recognise that this is an observational study based on lexicons. An appropriate next step would be to explore the use of word embeddings or semantic concepts to better capture the context in which these terms are used. Data mining techniques, such as clustering, could also be used to identify linguistic patterns that may emerge from the data. A study of influence in terms of users and messages for those communities, is also planned to explore the influence of language, how it spreads and from whom.

Despite these limitations, we have achieved our aim to combine feminist studies and social science models to make interdisciplinary observations about misogyny in the manosphere, over time and at scale. While we still have many different challenges to resolve, this work is one of the very first attempts to understand the manosphere both socially and computationally.

7 CONCLUSIONS

In this paper, we test existing feminist theories and models around the phenomenon of the manosphere by conducting a large-scale observational analysis of seven different online communities on Reddit, with openly misogynistic members. Grounded on feminist critiques of language, we created nine lexicons capturing specific misogynistic rhetoric and used these lexicons to explore how language evolves within misogynistic groups. Our results corroborate existing theories of feminist studies, particularly the idea that violence and hostility are increasing toward women online [30], that violent rhetoric and misogyny are co-occurring [8, 10], and and that stoicism and flipping the narrative are two contemporary responses to feminism [38].

REFERENCES

[1] [n. d.]. 2014 Isla Vista Killings. https://www.bbc.co.uk/news/world-us-canada-43892189. Accessed: 2019-01-27.
[2] [n. d.]. Alek Minassian. https://www.telegraph.co.uk/news/2018/04/24/do-know-alek-minassian-arrested-toronto-van-attack/. Accessed: 2019-01-27.
[3] [n. d.]. INCELS. http://haenfler.sites.grinnell.edu/subcultures-and-scenes/incels/#ffs-tabbed-311. Accessed: 2019-01-27.
[4] Maria Anzovino, Elisabetta Fersini, and Paolo Rosso. 2018. Automatic identification and classification of misogynistic language on twitter. In *Int. Conf. Applications of Natural Language to Information Systems*. Springer.
[5] Sarah Banet-Weiser and Kate M Miltner. 2016. # MasculinitySoFragile: culture, structure, and networked misogyny. *Feminist Media Studies* 16, 1 (2016), 171–174.
[6] Jamie Bartlett, Richard Norrie, Sofia Patel, Rebekka Rumpel, and Simon Wibberley. 2014. Misogyny on twitter. *Demos* (2014), 1–18.
[7] Lorraine Code. 1991. *What can she know?: feminist theory and the construction of knowledge.* Cornell University Press.
[8] Lorraine Code. 2002. *Encyclopedia of feminist theories.* Routledge.
[9] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009* (2017).
[10] Lynn Davies. 2008. Gender, education, extremism and security. *Compare* 38, 5 (2008), 611–625.
[11] Penelope Eckert and Sally McConnell-Ginet. 2013. *Language and gender.* Cambridge University Press.
[12] Elisabetta Fersini, Debora Nozza, and Paolo Rosso. 2018. Overview of the evalita 2018 task on automatic misogyny identification (ami). *Proc. 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA), Turin, Italy* (2018).
[13] Elisabetta Fersini, Paolo Rosso, and Maria Anzovino. 2018. Overview of the task on automatic misogyny identification at ibereval 2018. (2018).
[14] Michael Flood. 2004. Backlash: Angry men's movements. In *The battle and backlash rage on: Why feminism cannot be obsolete.* Xlibris Press, 261–278.
[15] Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 85.
[16] Rachael Fulper, Giovanni Luca Ciampaglia, Emilio Ferrara, Y Ahn, Alessandro Flammini, Filippo Menczer, Bryce Lewis, and Kehontas Rowe. 2014. Misogynistic language on Twitter and sexual violence. In *Proc. ACM Web Science Workshop on Computational Approaches to Social Modeling (ChASM)*.
[17] Russell G Geen and David Stonner. 1975. Primary associates to 20 verbs connoting violence. *Behavior Research Methods* 7, 4 (1975), 391–392.
[18] Debbie Ging. 2017. Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities* (2017), 1097184X17706401.
[19] Lise Gotell and Emily Dutton. 2016. Sexual violence in the 'manosphere': Antifeminist men's rights discourses on rape. *Int. J. Crime, Justice and Social Democracy* 5, 2 (2016), 65–80.
[20] Claire Hardaker and Mark McGlashan. 2016. "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics* 91 (2016).
[21] Sarah Hewitt, Thanassis Tiropanis, and Christian Bokhove. 2016. The problem of identifying misogynist language on Twitter (and other online social spaces). In *Proc. 8th ACM Conference on Web Science*. ACM, 333–335.
[22] Sylvia Jaki, Tom De Smedt, Maja Gwóźdź, Rudresh Panchal, Alexander Rossa, and Guy De Pauw. 2018. Online hatred of women in the Incels. me forum: Linguistic analysis and automatic detection. *Manuscript submitted* (2018).
[23] Marianne Janack. 2004. Feminist epistemology. Internet encyclopedia of philosophy.
[24] Emma Alice Jane. 2014. 'Back to the kitchen, cunt': speaking the unspeakable about online misogyny. *Continuum* 28, 4 (2014), 558–570.
[25] Emma A Jane. 2016. *Misogyny online: A short (and brutish) history.* Sage.
[26] Lori Kendall. 2002. *Hanging out in the virtual pub: Masculinities and relationships online.* Univ of California Press.
[27] Mary Lilly. 2016. *'The World is Not a Safe Place for Men': The Representational Politics of the Manosphere.* Ph.D. Dissertation. Université d'Ottawa/University of Ottawa.
[28] Jie Liang Lin. 2017. Antifeminism Online. MGTOW (Men Going Their Own Way). (2017).
[29] Kate Manne. 2017. *Down Girl: The Logic of Misogyny.* Oxford University Press.
[30] Alice E Marwick and Robyn Caplan. 2018. Drinking male tears: language, the manosphere, and networked harassment. *Feminist Media Studies* 18, 4 (2018).
[31] Adrienne Massanari. 2017. # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society* 19, 3 (2017), 329–346.
[32] Angela Nagle. 2017. *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right.* John Hunt Publishing.
[33] Mohammadreza Rezvan, Saeedeh Shekarpour, Lakshika Balasuriya, Krishnaprasad Thirunarayan, Valerie Shalin, and Amit Sheth. 2018. Publishing a Quality Context-aware Annotated Corpus and Lexicon for Harassment Research. *arXiv preprint arXiv:1802.09416* (2018).
[34] Katy Sian. 2018. Stupid Paki Loving Bitch: The Politics of Online Islamophobia and Misogyny. In *Media, Crime and Racism*. Springer, 117–138.
[35] Kat Stoeffel. 2014. Women pay the price for the Internet's culture of anonymity. *The Cut* 12 (2014).
[36] Jenny Sundén and Susanna Paasonen. 2018. Shameless hags and tolerance whores: feminist resistance and the affective circuits of online hate. *Feminist Media Studies* 18, 4 (2018), 643–656.
[37] Joewon Yoon. 2018. Down Girl: The Logic of Misogyny. *Asian Women* 34, 1 (2018), 109–112.
[38] Donna Zuckerberg. 2018. *Not All Dead White Men: Classics and Misogyny in the Digital Age.* Harvard University Press.