



Article

Topic Modeling The Red Pill

J. B. Mountford

Independent Researcher, Brighton BN1 4SG, UK; jbm112358@gmail.com

Received: 10 December 2017; Accepted: 1 March 2018; Published: 9 March 2018

Abstract: The Men's Rights Activism (MRA) movement and its sub-movement The Red Pill (TRP), has flourished online, offering support and advice to men who feel their masculinity is being challenged by societal shifts. Whilst some insightful studies have been carried out, the small samples analysed by researchers limits the scope of studies, which is small compared to the large amounts of data that TRP produces. By extracting a significant quantity of content from a prominent MRA website, ReturnOfKings.com (RoK), whose creator is one of the most prominent figures in the manosphere and who has been featured in multiple studies. Research already completed can be expanded upon with topic modelling and neural networked machine learning, computational analysis that is proposed to augment methodologies of open coding by automatically and unbiasedly analysing conceptual clusters. The successes and limitations of this computational methodology shed light on its further uses in sociological research and has answered the question: What can topic modeling demonstrate about the men's rights activism movement's prescriptive masculinity? This methodology not only proved that it could replicate the results of a previous study, but also delivered insights into an increasingly political focus within TRP, and deeper perspectives into the concepts identified within the movement.

Keywords: machine learning; digital humanities; men's rights activism; masculinities; online communities

1. Introduction

The internet permits more taboo and extreme expressions through anonymity (Lyons 2017) and post-geographical connectedness (Sardar 1995). Therefore, many of the most extreme and insightful new gender performances are observed online before being identified offline. Furthermore, given the vast scale of textual data that is produced by users online everyday (estimated to be 2.5 quintillion bytes daily in 2013 (Wu et al. 2014)), the resources required to process that data in open and closed coding practices become infeasible, due to the quantity of content the coder has to process. Natural Language Processing (NLP) methodologies are the solution to this problem. Using Latent Dirichlet allocation (LDA) (Deerwester et al. 1990) and Word2vec (Mikolov et al. 2013), topics within documents can be grouped and words can be queried to find similarities to other words, which can give insights into the author's overall views without having to read a significant portion of their corpus. This methodology could be extended by taking into account more sources, including online forums (Mountford 2015) and other websites, such as those listed in Schmitz and Kazyak (2016), and assertions could be made about shifts over time, the movement's methodology and more.

Linked to the increasingly influential Alt-Right (a coalition of far right, traditionalist Christian, disenfranchised geeks, and pickup artists (Lyons 2017; Kelly 2017)), and part of the anti-feminist online movement called the manosphere (Ging 2017), the Men's Rights Activism (MRA) movement and The Red Pill (TRP) are online groups creating novel performances of hegemonic masculinity (Connell and Messerschmidt 2005). Hegemonic masculinity is a performance of masculinity that aims to maintain a dominant position for men typified by subordinating women and non-traditional performances of masculinity through a variety of techniques, including violence, societal structure, and discrimination

(Connell and Messerschmidt 2005). TRP is a group based around a philosophy that rejects modern feminism and progressivism, believing instead in a genetically deterministic conception of gender: it asserts that genders have inherent roles due to their physical characteristics (Beynon 2001; Ging 2017). The movement originated in 2009 (Ging 2017) and is centred around online discussion boards, with core personalities and thinkers using blogs and vlogs as platforms for distributing their thinking.

The potentials of natural language processing (NLP) to infer topics, sentiment, and more from naturally produced text, to yield insights without the researcher having to read through the content, are significant (Jacobi et al. 2016). Natural language processing is a field of research that aims to allow computers to process text and to identify the same meanings, subjects, and associations as a human would (Forsyth and Martell 2007). The use of NLP data models will allow a degree of reproducibility and reliability that coding methodology performed by researchers, while being highly insightful and useful, is unable to offer, with manual coding liable to being influenced by the coders and limited by the scale of the data that can be coded (Jacobi et al. 2016). This paper demonstrates that faster and more robust insights can be made about a community by using a combination of two data models—one for fine detail in small scale, and one with a wider scope and larger scale. The trained models can then be shared in online code repositories, and can be reused by other researchers for comparison, combination, and calibration of further data sets. See Supplementary Materials.

The NLP data models this paper will use are Latent Dirichlet Allocation (LDA) (Deerwester et al. 1990) and Word2vec (Mikolov et al. 2013). LDA (Deerwester et al. 1990) is a data model that assumes that if words are found in the same context, then the words are semantically similar. Based on the words found in documents, LDA searches for underlying similarities in term usage within documents, to isolate patterns of similarity that form semantic topics (Deerwester et al. 1990). The Word2vec data model (Mikolov et al. 2013) builds a word's similarity based on the terms that surround it every time it is used. The Word2vec model then looks for terms with similar surrounding words, and will return these as semantically similar to the original (Mikolov et al. 2013). Neither model requires priming with data that has been translated by a researcher, to "teach" it how to interpret, making both models "unsupervised algorithms" (Mikolov et al. 2013; Deerwester et al. 1990), which will be free from researcher bias in how they assemble semantic similarities; the models can only base their results on the text they ingest.

TRP is a community focused on creating a masculinity to counter what it believes is an insidious feminist campaign to dismantle traditional masculinity, and to intentionally harm men (Kelly 2017). By recommending or prescribing their audience to perform this masculinity through lifestyle and self-help content, TRP and Return of Kings (RoK) can be seen as prescribing a new form of masculinity that is consciously constructed in reaction to feminist and societal shifts; this is the prescribed masculinity that this paper seeks to understand. This masculinity is orientated to achieve the traditional hegemonic aims of sexual conquest, social dominance, and self-improvement using misogynistic philosophy (Gotell and Dutton 2016; Connell and Messerschmidt 2005), which delegitimises feminism's arguments through anecdotal rebuttals (Ging 2017). There is significant insight potential, as this new development shows the directions and methods that anti-feminist masculinities are likely to develop in the future.

As TRP is predominantly an online movement, making up part of the overwhelming avalanche of written data uploaded to blogs, social networks, and message boards, the quantity of data created is both a massive opportunity and a colossal challenge to analyse. The most recent and in-depth study of the community, "Masculinities in Cyberspace" (Schmitz and Kazyak 2016) initially used open coding, followed by closed coding methodology, but was limited to fifty articles from twelve sites distributed over twelve months from which to draw its insights. This paper aims to add to this work by diving deeper into one of the sites studied, Return of Kings (returnofkings.com), using topic modeling methodology to attempt to replicate the coded themes identified in "Masculinities in Cyberspace". Return Of Kings (RoK) is a blog, run by Daryush Valizadeh under the pseudonym Roosh V. RoK aims to "usher the return of the masculine man" with the caveat that "yesterday's masculinity is today's

misogyny". Roosh is a writer and coach on men on how to pick up women. RoK forms a central voice in TRP, and writes on topics from political trends (Luthra 2017) to legalised rape (Valizadeh 2015), but mostly focuses on arguments that western society has failed men.

Using this methodology, this paper demonstrates what insights topic modeling can provide into TRP's prescriptive masculinity. To prove that the methods are reliable, this paper compares the topics produced by the data models against the topics manually created by Schmitz and Kazyak. Furthermore, this paper aims to analyse the coherence of these topics manually, to assess the insight potential of these clusters. By automating as much of the research as possible, this paper demonstrates the practical time and resource-saving potential of these methodologies, balanced against the limitations of clarity and depth.

2. Literature Review

MRA and TRP can trace their lineage from anti-feminist reactionaries (Messner 2016), via a subversion of gender studies terminology and theory, such as the crisis of masculinity and the feminist mystique (Messner 2016)—for example, proposing that men suffer from gender roles more than women (Gotell and Dutton 2016). This perspective was bolstered by the rise of post-feminism and public views that feminism had achieved its aims (Lyons 2017). Disenfranchised, economically unstable, blue-collar men were increasingly affected by the liberal and progressive societal changes (Messner 2016). Traditionally, these men would have performed a class-relevant form of hegemonic masculinity (Connell and Messerschmidt 2005), constituted by homophobia, objectification of women, and subordination of others, or a masculinity that referenced the power that hegemonic masculinity wielded.

Men turned to movements to give them answers on how to perform masculinity (Kelly 2017). The seduction community was one of the first neo-masculinities that scratched this itch, with ascetic hedonistic teachings and promises of being more attractive to others, and ultimately self-improvement (Hendriks 2012). Ascetic hedonism is the mindset and methodology used by pickup artist mentors, which teaches that in order for students to achieve their goals of sexual conquest, they must reject this aim, and focus instead on the skills required to achieve this aim: a paradox of ascetic rejecting of aiming to have sex with women, and hedonistic aiming to improve the self to become more attractive to women. Hendricks noted the complex discipline of self that led to the objectification of both self and others. The motivation of this discipline was split initially between the male ideal of sexual relationships and self-development.

TRP can be seen as a product of the online space; the unique formation and spread of the movement is based on inherent characteristics of the internet. The online space has been extensively studied for gender bending (playing characters of different genders (Kendall 2002)), challenges to societal norms of gender (Mowlabocus 2006), and as a space for gender formation (Light 2013; Hendriks 2012). It facilitates a common ground for users to gather and collectively explore gender (Biricik and Hearn 2009). These features allow for TRP to develop a gender identity that is strongly counter-cultural, without concern for physical distance or their taboo views.

Research on the online space has identified several core characteristics, with the most relevant to this research being its provision of an anonymous space for some users to explore their performance of gender, and to consume content related to gender (Johnson 1997; Boyd 2008). This anonymity, inherent to the online space, permits users with non-socially-acceptable views to find others and strengthen their identity (Mowlabocus 2006), as in the case of gay men seeking high-risk sex (Dowsett et al. 2008) and in the case of TRP's anti-feminist, right-wing views. The disruption of geography that permits a usually-dispersed extreme minority to congregate, combined with a silent majority of lurkers (Van Dijck 2009), facilitates a form of growth that is typified by extreme views permeating, without a context of moderation that might be expected in offline spaces. In this way, the extremeness of TRP's views is not unexpected, but would likely be impossible without the internet.

TRP is expanding its membership online using the innate features of the internet, through a relational nature with feminist outrage. Gotell describes the reactionary nature whereby critical feminist attention amplifies their message (Gotell and Dutton 2016), potentially because of the TRP technique of presenting victimhood at the hands of feminism. Furthermore, the anonymous nature of the internet can facilitate more extreme representations of offline gender performances (Lyons 2017), which would amplify and attract those seeking to solidify a gender that is increasingly culturally unacceptable. Finally, the viral spread of TRP can be linked to the reproducibility and searchability of the content that is produced (Johnson 1997); where traditionally the collective gender identity would have to spread through physical contact, the internet allows content to spread independent of its creator, through social networks and content aggregation hubs. The speed of this viral expansion is one of the reasons why fast, automatic research techniques are useful, as they increase the speed in which insights into motivation and core concepts can be identified.

TRP aims to build a prescriptive masculinity that can take advantage of this worldview to maximise its audiences' prescribed aims of sexual relationships and self-growth (Schmitz and Kazyak 2016). These aims imply a masculinity that is looking for answers on how to deal with the changing expectations of men, within both the crisis of masculinity and the rejection of traditional masculinity. In this way, men seeking guidance on how to construct their masculinity, because traditional role models have been called into question (Schmitz and Kazyak 2016), are drawn to strong characters and the provided proof and the reassurance of male superiority that the manosphere represents. The manosphere is one of the alternative masculinities referenced in Light (2013) that is facilitated through the anonymity of the internet.

In their analysis of the MRA and TRP communities, asking the question "what strategies do online MRA groups utilize to delegitimize feminism and the goals of gender equality?", Schmitz and Kazyak (2016) found two meta-themes: "Cyberlads in search of masculinity" and "Virtual Victims in search of equality", containing three and five subthemes, respectively. These subthemes comprise the method techniques that the studied websites use to increase their legitimacy, power, and influence within the MRA group. The strategy that Cyberlads utilize is creating hyper-masculine, lad-culture-influenced lifestyle content that empowers men. The sub-themes of Cyberlads are the homosocial policing of masculinity through self-help and lifestyle content, espousing the evils of feminism through "myth busting" and portraying women as sexual commodities, by positioning them as the objects of success through pickup. Virtual Victims approach their aims differently. The sub-themes within Virtual Victims are portraying men in crisis, combating institutional misandry and delegitimizing women's issues. By repurposing sociological language and theories, MRA writers position men as discriminated against by society, requiring activism to empower men.

The significant presence of political topics in the RoK data models is not unexpected, as the rise of populism and right-wing views has impacted TRP due to the similarities of their positions on inherent equality (Lyons 2017). MRA and TRP's rejection of gender equality found resonance in the alternative right (Alt-Right), which is united by a rejection of human equality (Lyons 2017) in the philosophy of Identitarianism. Lyons notes the affinity of TRP's manosphere with the non-equalitarian philosophies uniting the Alt-Right, and how the manosphere is viewed as a potential entry for men disaffected by the changing political landscape, causing the crisis of masculinity to enter into the Alt-Right's views. This, combined with the increase in traditionalist conservative discourses in media gives context to the political topics that are shown in the data models.

Topic modeling and NLP research methods have been recommended for use within digital humanities and social sciences for some time (Conte et al. 2012; Leonard 2014), but the skills required for these methods has restricted researchers from taking full advantage of the benefits. These benefits are shown in the ability to analyse very large corpora to yield insights into topic groupings (Jacobi et al. 2016). Jacobi et al. base their methodology on expanding the existing coding and analysis of "frames" or topics by Gamson and Modigliani (1989), asking the question whether topic modelling could replicate Gamson and Modigliani's findings. Jacobi et al. do this by building LDA topic models

optimised against a perplexity curve, then mapping the frequency of these topics being identified against the date of each article, to create a time-series chart. Using a similar methodology, it is shown that LDA topic modeling can be used to yield insights about community identity, gender construction, and the persuasive discourse used to justify the extreme counter-cultural masculinity models described by [Schmitz and Kazyak \(2016\)](#).

3. Methodology

The online space is ideal for purely text-based NLP topic modelling analysis, as many paralinguistic features of communication are stripped out by the inherent nature of the medium. It would be incorrect to suggest that the proposed methodology would be appropriate to analyse transcripts of pickup artist coaching, as the body language and the wider context of rapport between the participants might be more meaningful than the actual words said. However, in the context of blogs online, the audience is only reading the blog articles, removing the potential of rapport and past history to corrupt NLP analysis. This reduction in communication features outside of text allows us to be more confident in the accuracy of the two model's outputs, through the very nature of the internet.

Return of Kings was chosen for two reasons. Firstly, it is acknowledged as an important site to the manosphere, and has been extensively analysed in multiple studies ([Ging 2017](#); [Schmitz and Kazyak 2016](#); [Lyons 2017](#); [Southern Poverty Law Center 2012](#)). This provides an existing analysis to compare against the methodology proposed in this paper. Furthermore, the author of the site, Roosh V, is one of the most well-known members of the manosphere, being called in an interview "The web's most infamous misogynist" ([Price 2015](#)). Secondly, the site was able to be crawled cleanly, with a minimal amount of HTML found within documents, which would distort the content groupings that rely on relevant topic words being found next to each other. If the model was to mistake code that should be invisible to readers for topic-relevant information, it would distort the topic training process.

To crawl RoK, a web spider built in the Python library BeautifulSoup ([Richardson 2013](#)) was used to download pages, extract the text and internal links from the downloaded HTML, and then to add those internal links to the list of pages to download. The text from these pages was then segmented into paragraphs, to improve the coherence of the document/topic distribution, and all duplicate documents were then removed, to avoid default headers and footers being overrepresented. All blog posts on RoK were crawled, resulting in a total of 4983 documents after the URLs were filtered to only leave the blog posts themselves. All of these documents were fed into the next stage of processing, removing any need for sampling affordances. There is no theoretical limit to the number of documents that can be included in this method of analysis, as the relevance and term-term or term-document associations are stored at the cost of more intensive, slower, and expensive computing requirements.

To prepare the data, the documents were split using a lemmatizer, and then filtered by parts of speech to leave only nouns, in order to improve performance and human coherence using the Natural Language ToolKit ([Bird et al. 2009](#)). The lemmatizer simplified variants of word roots to optimise the LDA's document term matrix ([Tirunillai and Tellis 2014](#); [Khandelwal and Harada 2004](#)). This left a sanitised corpus of 59,094 documents, with a total of 2,150,151 words. This sanitised corpus was modeled into an LDA and a Word2vec data model, using the Python library Gensim ([Rehurek and Sojka 2010](#)). For future usage by researchers, these datasets and models were made available in an online corpus repository ([Mountford 2018a](#)), with the analysis available in Python notebooks ([Mountford 2018b](#)).

The number of topics was decided by optimising for perplexity and coherence within the LDA topics, following Jacobi et al.'s methodology ([Jacobi et al. 2016](#)). Fourteen topics were chosen, on the basis that they showed the highest levels of coherence and lowest levels of perplexity, within a total number that a researcher could be expected to give meaningful titles to; topic sizes between five and forty were tested. Then, using a topic browser similar to that described in [Schmitz and Kazyak \(2016\)](#), the topics were titled based on the most significant terms and the documents that had the highest levels of affinity for these topics.

The topics that were formed are coherent with any HTML that was inline in the text being isolated from the topics this paper will study. Because of the repetitive nature and limited vocabulary of HTML, this code been isolated into its own topic. This has been labeled “Navbar”, as the most representative terms were those that were used in the navigation bar and are ignored from analysis.

These topics were then reapplied to the documents the topics were trained on, in order to analyse the patterns within which the topics would be found within the same document, implying that the subject matter and motivations overlapped. From this, an overlap matrix could be created, using the average percentage certainty with which the model predicted each topic to be in each document, using a phi value probability.

These overlaps were then applied to a co-occurrence chart, with a cutoff that removed all co-occurrence links below 0.03 to yield the most definition. This filtered co-occurrence matrix was then turned into a network graph (Figure 1), with topics as nodes and edges representing the co-occurrence strength. The colours of the nodes were set based on the number of edges connecting it to others, and the colour of edges was based on the overlap from the phi value probability. The purpose of this graph was to visually understand the co-occurrence network, in terms of what topics were most linked to other topics (hubs), where clusters of topics formed, what topics were only co-occurring with only a few other topics (outliers), where the strongest co-occurrences were (signified by stronger edges), and what topics feature such weak co-occurrences that they would not be represented at all.

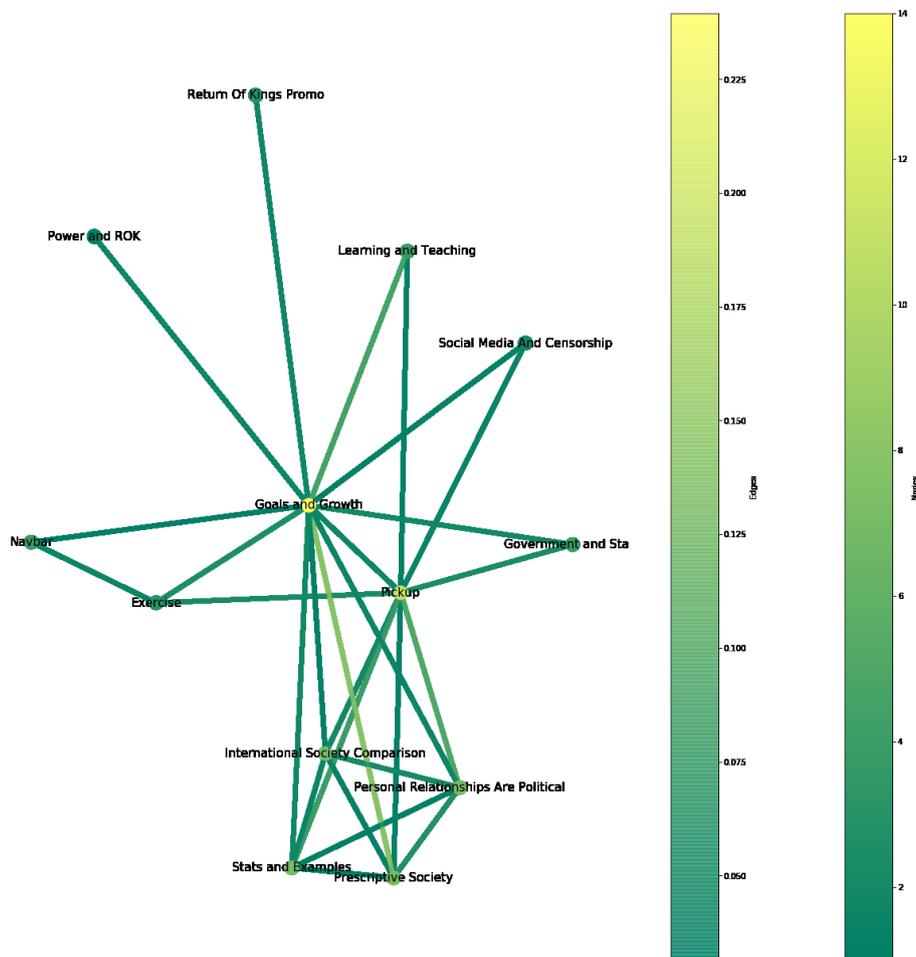


Figure 1. Network graph of topic co-occurrence. Edges are coloured by co-occurrence strength. Nodes are coloured by the number of links remaining after the removal of all edges under a strength of 0.03.

Due to the term-to-term comparison limitations in the LDA model, for deeper, ad-hoc analysis a Word2vec model (Mikolov et al. 2013) was trained on the same sanitised corpus and dictionary as the LDA model, using default parameters. With this model, it's possible to quantify words used in similar contexts and words that have similar collocations. This allows the researcher to drill deeper into the usage of certain words. Similarity scores between topics are calculated from cosine similarity of the vectors.

For example, if the model registers the word "princess" within the corpus as highly representative of the topic, the Word2vec model allows the researcher to drill into the links between "princess" and the terms "Disney" (with a similarity of 0.85), "spoilt", and "Cinderella", whereas the word "prince" has more historical, spiritual linked words: "biblical", "prophet" and "everlasting" (Mountford 2018b). From this, the researcher is able to see that female positions of power are only written of in a fictional sense, whereas male positions of power have historical and religious weight attributed to them.

Benefits of This Methodology

The depth of the insight available to the coder in Schmitz and Kazyak (2016) was limited to the number of articles they were able to ingest, code, and analyse. With more data, is it unreasonable to propose that more insights can be extracted from how members of TRP write and therefore what they believe? This paper believes that the LDA/Word2vec methodology is able to provide coherent insights, and that it can be applied to a similar if not larger number of sources, to find more reliable insights into how TRP and other communities interact. By limiting the scope of this initial paper to only one source, the accuracy can be proven without time investment into the crawling and processing of many sources. Furthermore, by keeping the source scope small, this paper is able to build on the less in-depth analysis of textual analysis in the same field, such as Jane (2017), by analysing the context of language used instead of the more limited data model of Jane's "Random Rape Threat Generator". This would achieve the same aims as Jane of understanding how TRP can be interacted with, despite the counter-productive vicious cycle that Gotell and Dutton (2016) identify, through understanding the relational nature of feminist outrage and manosphere victimhood.

Ging (2017) performed a similar research project to that proposed in this paper, by categorising and topic clustering in the manosphere, but was limited to only 38 sites and one item of content for analysis in its critical discourse analysis. Regardless of the reliability of Ging's sampling method of intuitively identifying interlinked sites, the methodology proposed in this paper allows for whole population analysis, as the models are built on the content from the entire available corpus. This methodology avoids typical sampling errors in non-random samples, as well as ascertainment bias, where important representations of a trend can be ignored after the symptoms of the trend are excluded from analysis.

4. Results

Analysing the frequency with which the topics were identified (as graphed in Figure 2), there are three obvious groupings. "Goals and Growth" is by far the most common classification, being found in 46% of documents.

The core topics of "Pickup", "Personal Relationships Are Political", "Prescriptive Society", "Stats and Examples", and "International Society Comparison" are the second grouping, with frequencies between 28,647 and 24,313, being found in 7965 more documents than the next closest grouping.

Between frequencies of 16,348 and 10,071 are the complementary topics of "Teaching and Learning", "Government and State", "Social Media Censorship", and "Exercise", in that order. The prevalence of lifestyle content, both in the topics within RoK and the style of the content written by Cyberlads, indicates strength in the reproductive powers of this methodology.

Investigating the network graph of topics (Figure 1), a similar distribution to that of the frequency graph (Figure 2) can be seen. There are two hub topics of "Goals and Growth" and "Pickup", with a cluster of topics that are both linked to the hub topics and closely to each other. Beyond these densely interconnected groups there are outlier topics that are only linked to the two hub topics.

This distribution suggests that RoK has two hub topics that are always referenced in content, as they are central to the site’s ideological positioning, complimented by special outlier topics that serve to illustrate and augment these hubs. The distribution also suggests that there is a second group of topics that, whilst not as central as the hub topics, is usually mentioned with each other, implying a repetitive nature to the way that they are discussed.

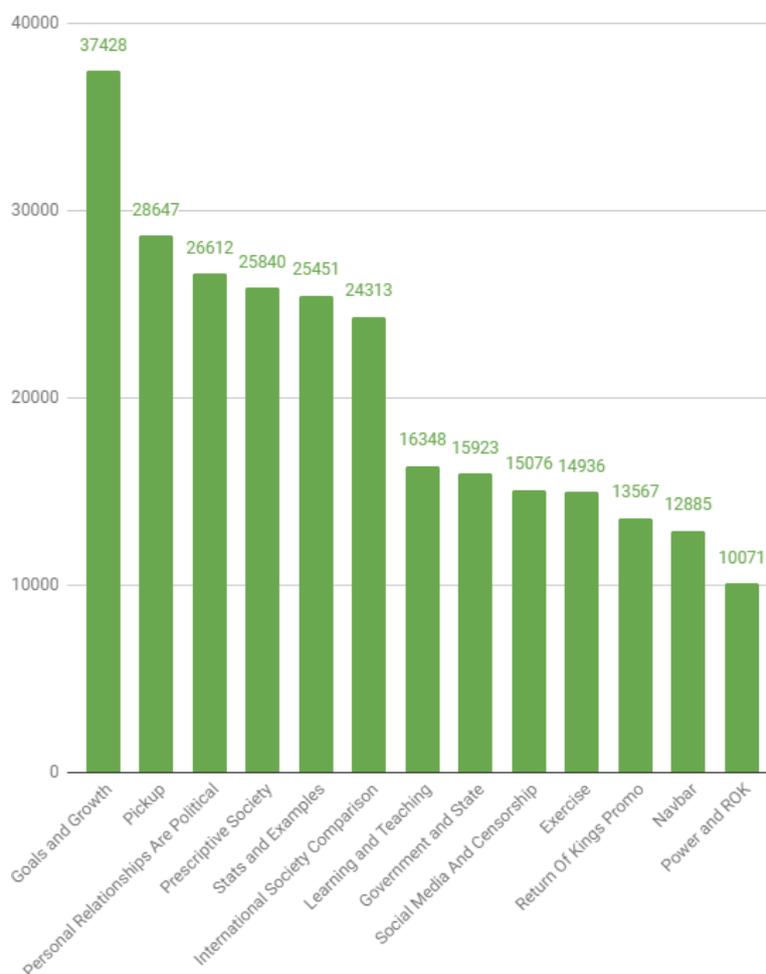


Figure 2. Frequency chart of how often topics were identified within the corpus.

4.1. Self-Help, Lifestyles and Ascetic Hedonism

4.1.1. Goals and Growth

From the topics that were formed from the model, the most prevalent by far is that of “Goals and Growth”, being found in 46% of all documents, 8781 more documents than the second largest topic. The topic was named as such because of its most characteristic words being the aspirational “get”, “life”, and “time”, and because the most representative document of the topic regarding self-improvement to achieve success was about the desire for success, the acceptance of suffering justified by the improvement of technique, and the importance of learning (Mountford 2018b).

The “Goals and Growth” topic serves as one of the two core hubs underlining the deep link of self-improvement content and prescriptive masculinity as a solution to RoK’s audience’s uncertain aims regarding gender identity. “Goals and Growth” is the reason why audiences are reading the website: the topic is the unique selling point of RoK. TRP aims to help men understand their place in a world that is rejecting hegemonic masculinity (Connell and Messerschmidt 2005). If men are

disempowered because of the decline of the patriarchy, or lonely because of shifting societal gender roles, the “Goals and Growth” hub is the solution to this.

The Word2vec connotations of the words typifying the “Goals and Growth” topic back up these conclusions. “Goals” links to ascetic terms, such as “discipline”, “mindset”, and “strive” (Mountford 2018b), and less to the reason for pursuing the goal in the first place. Drilling into what achieving the goal would look like, the word “success” links to physical rewards, terms like “valuable”, “reward”, and “gain” (Mountford 2018b); however, “success” also links strongly to the ascetic process of sacrifice, with the terms “discipline”, “dedication”, “motivation”, and “effort” (Mountford 2018b). These insights support Hendriks’ (Hendriks 2012) propositions that discipline itself becomes the goal in ascetic hedonism.

Whilst only found in 20% of documents, the topic “Learning and Teaching” has a strong link to “Goals and Growth”, despite being an outlying hub. Characterised by words like “book”, “help”, and “intellectual”, and by topics talking about “dissident intellectuals”, the topic is certainly not rejecting education or intellectualism; however, it is rejecting other ideological groups. For example, “cuckservatism” rejects moderate conservatism, and “victim” “leftist[s]” rejects left wing politics’ aims of equality against the victims of discrimination (Mountford 2018b). Being the seventh most mentioned topic, it is mentioned in only 20% of documents, but shows a strong overlap with “Goals and Growth”, in a similar way that “Pickup” is strongly linked to “Exercise”. The inferences of this are that while Exercise is a topic that is important to TRP, intellectual development is still valued, and indeed it can form a necessary part of the goals that TRP recommends striving for, provided that it does not involve researching groups that question or go against its teachings.

4.1.2. Pickup

The second most frequent topic, and also the second most-interlinked hub topic, regards pickup. The pickup topic, found in 35% of documents, is focused on prescriptive dating advice and recounting encounters with women, who are predominantly referred to by their physical appearance or gendered insults, e.g., “bitch”, “blonde”, and “brunette” (Mountford 2018b). This maps against the two approaches that The Red Pill use for developing their arguments according to Schmitz and Kazayak (2016); these are a lifestyle prescription and worldview framing that offer “a breadth of lifestyle advice aimed at empowering men and encouraging them to unapologetically embrace their masculinity”. The topic “Pickup” is clearly made up of advice and examples. The most representative words of the topic are “get”, “just”, and “go”, and the most representative documents encourage the reader to “imagine” being in a situation. The topic also includes recollections of Roosh’s experiences about times he has had sexual success with women. The second most-representative document refers to the women only by their hair colour, and documents their and Roosh’s actions as part of the wider use of learning examples as a teaching tool, with the writer giving strategies and methods through evidence-based examples.

The portrayal of women as sexual commodities in the topics generated by the model comes in two flavours: overt in the language in “Pickup”, and implicit in the case of “Goals and Growth”. Overtly, women are referred to as identifying features, rather than as people; they are blondes and brunettes that are linked to further objectification through their connotations with “wearing”, “tit”, “bra”, and “dressed” (Mountford 2018b). Furthermore, the use of “girl” as the female collective noun compared to the male “guy”, in the terms most representative of the pickup topic, demonstrates the gendered and sexist language used. Whilst these insights are neither new nor unexpected, the fact that the data models have isolated it as a trend indicates the models have reliability in replicating existing theory.

Drilling deeper into the connotations of “girl”, compared to other female pronouns of “woman”, “lady”, and “female”, significant differences can be seen (Mountford 2018b). “Girl” has strong links to “guy” and “hot”, but also with negative and promiscuous terms, such as “slut” and “slutty”, indicating that girls are predominantly an object of sexual desire. The other female pronouns for comparison all share the links to promiscuity, but differ in terms of the wider perception of the term. “Woman” has

connotations with “pedestal” and “entitlement”, linking to the TRP belief that feminism has elevated women to a level they do not deserve. “Lady” is more linked to weddings and long-term partnership, with terms such as “dress”, “girlfriend”, and “sweet”. “Female”, being a more scientific term, has links to more theoretical terms, like “decriminalized”, “perpetuating”, “bolstering” and “devaluing”.

The common links to promiscuity maps against Schmitz and Kazyak’s identification of the Madonna/whore dichotomy in MRA writings, whereby women are both presented as vulgar objects of sexual desire and also prescribed to remain pure, virgin, sexless beings. Looking into the connotations of the word “virgin”, this trend is seen to be borne out. Virgins are mentioned in similar contexts as marriage (“marry”, “marrying”, “bride”) and as long-term partners (“loyal”, “girlfriend”), but are also linked to promiscuity (“slutty”, “whore”, and “debaucherous”). Again, the models’ ability to replicate this trend proves their reliability as research tools.

4.1.3. Political and Societal Gender

Beyond these two core hub topics, there is the nest of highly interlinked topics: “Stats and Examples”, “Personal Relationships Are Political”, “International Society Comparison”, and “Prescriptive Society”. These all share a descriptive element, interpreting and reflecting societal and political shifts to re-enforce the MRA viewpoint, and are all present in 30–32% of documents. “Stats and Examples” is self-explanatory, using selective stats and cherry-picked examples as it plays on the inherent “reproducibility” of the internet (Johnson 1997), whereby users can link to and reference information without having to put it into context. The focus of these ancillary topics on political or societal shifts suggests they are seen as a method of pursuing the aims expressed in the core topics, rather than core topics themselves. This resonates with the Alt-Right’s positioning of Donald Trump as a method for shifting the window of permitted political positions, rather than seeing Trump as a member of their movement (Lyons 2017). Furthermore, the repetitive nature of this interlinked nest’s topic usage would map against the expected repetitive nature of political content, which is likely to always cover the same ground.

Two similar topics in the cluster of interlinked topics are “International Society Comparison” and “Prescriptive Society”, which are closely linked to the overarching societal or cultural perspective that RoK uses to criticise and prescribe an ideal society. These topics, sixth and fourth most common within the corpus, respectively, have a comparative element, with “International Society Comparison” referencing Chinese, Russian, and Middle-Eastern politics and politicians, and “Prescriptive Society” mentioning and discussing the changing and perceived-to-be more prescribed views on transgenderism, religion, and progressive politics. The dynamic between the two topics goes beyond their titles, as “International Society Comparison” takes into account the election of Donald Trump, with the most characteristic term being “Trump”, echoing the representative terms “white”, “group”, and “world” in “Prescriptive Society” (Mountford 2018b). The importance of comparison between the political and societal positions indicates the search for the ideal masculinity or culture by TRP, often rejecting the western model because of its perceived dominance by feminist thinking.

The overlap edge of “Prescriptive Society” with “Goals and Growth” is also the strongest of all topic overlaps, indicating that the two topics are found alongside one another around 16% of the time. This links to one of the most interesting points: how can RoK both be prescriptive in how it tells its members to behave, whilst also rejecting and complaining about society by prescribing changes to it? The link between “Prescriptive Society” and “International Society Comparisons” is the answer to this question; it is not a prescription when it is framed as a return to a traditional practice. Indeed, the historical leaders that are presented in “International Society Comparison” are not only presented as societal leaders, but as political figures, and also as personal figures for the audience to learn from. The difference for RoK is between a prescription of a progressive change that has not yet taken place, with effects that are not yet known, and in which traditionally privileged men are finding women elevated to their equals, as opposed to a traditional way of life, which is presented as successful and the basis for current societal success.

The final topic in the interlinked cluster is that of “Personal Relationships Are Political”. This topic is more nuanced, but is named because “sexual”, “sex”, “woman”, and “men” are the most characteristic words, with the most typical documents showing far greater venom and hatred than other topics: “men”, “time”, “turned”, “weaking”, “proud”, “mesmerized”, “lipstick-wearing”, “bitch”, “men”, “dependent”, “spoiled”, “woman”, etc. Notable is the use of “woman” and “man”, with the diminutive “girl” in Pickup replaced with the more adult “woman”, signalling that content in this topic is written in more abstract tones of general genders, rather than concrete examples (Mountford 2018b). This topic shows the interpretation of romantic relationships through TRP’s perspective as stories of manipulation and power. This is potentially an extension of the de-personalising and de-emotionalising processes described in the seduction communities’ interaction with women, as described in Hendriks (2012). Drilling into the similar terms as these representative terms, “proud” shows links to mostly negative terms, implying that even positive terms such as this are linked to features of hatred, like “spoiled” and “princess” (Mountford 2018b).

“Princess” is a frequently used term within the corpus, and is a central concept to the modeled topics. It is used to infantilize women, with similarities to “Disney”, “brat”, and “doll”, whilst also sexualized with terms such as “slutty” and “whore”, which is seen with all terms referring to women (Mountford 2018b). Interestingly, there are also links to strong male characters; Robin Thicke and John Travolta both show up with links to “suave” and “stylish”. If the vectors of the promiscuity terms are subtracted from the vector of “princess”, the returned terms only refer to powerful men, implying both that the promiscuity vector is very similar to the female vector, but also that “princess” does imply a degree of power. The most similar terms returned to this query are male positions of power, such as “emperor”, “statesman”, “Julius”, suggesting that while “princess” is used in a diminutive sense, there is an inherent power to the position, which is tied up with “princesses” acting as “bratty” and “spoiled” through “betas” that pander to their supposed brattyness.

4.1.4. Outlier Topics

The notable outlier topics (topics that are only connected to the two core hubs) are “Social Media and Censorship”, “Government and State”, “Teaching and Learning” and “Exercise”. “Exercise” is notable in its lack of connection to “Personal Relationships Are Political”, implying that RoK rarely writes of exercise as a means for achieving long-term relationships; rather, it is only for pickup and short-term relationships. Furthermore, the topic not linked strongly to societal or political topics. Exercise is purely personal. It yields social rewards by making the practitioner more attractive, but is not linked strongly to the more personal “Goals and Growth”. The exercise is to demonstrate strength to women.

The “Social Media and Censorship” topic (found in 18% of documents) shows the projected victimisation of TRP, which sees itself as censored by the feminist media that is restricting what views are acceptable online and offline. The most representative documents are those exploring events on social media, including the resignation of Milo Yiannopoulos from Breitbart, Yiannopoulos having his twitter account removed by Twitter moderators, and the rise of Medium as a left-leaning crowd-blogging platform. The importance of social media to TRP is high—indeed, it is one of the core mediums of the movement, allowing users to find others that may be far away, but who are equally passionate. These interactions would not happen pre-internet, and they are facilitated by the ease of content creation and of finding users and groups of similar mindsets to the reader (Johnson 1997). The fact that social media has a topic to itself implies that it is both a medium and a subject to TRP, which views it as a means to an end of disseminating their message as well as a battleground where they are restricted from their aims.

4.1.5. Virtual Victims and the Evils of Feminism

Schmitz and Kazyak identify a methodology of MRA called “the Evils of Feminism”, where MRA finds incoherencies or injustices performed by feminists or those who have been influenced by feminists.

While this methodology can be found in this papers corpus, it is not a distinct topic; rather, it is a theme that runs through all topics. RoK positions the political left as the insidious cause of the degeneration of society (Kelly 2017). It is no coincidence that one of the key documents typifying “Prescriptive Society” refers to leftist double standards, echoing the use of feminist incoherencies, or “myth-busting”, as Schmitz and Kazyak (2016) describe it. The construction of the internet, with unlimited reproduction and re-interpretation, facilitates myth-busting as one of the most persuasive attacks that can be made by TRP.

Furthermore, the fact that Schmitz and Kazyak’s (2016) method of “Virtual Victims” is not a coherent, independent topic may be insightful, as the construction of the Virtual Victim is difficult to reconcile with a position of power. If RoK is projecting a powerful masculinity, dominating both race and gender, it would undermine that projection to admit that it is simultaneously dominated by government, society, and society’s expectations. When searching for similar terms in the corpus with the Word2vec model, it can be seen that the term “victim” is linked with false accusations, with terms like “innocent”, “blaming”, “purported”, and “portrayed” (Mountford 2018b), linking to the legal discrimination that Schmitz and Kazyak (2016) identify as an MRA talking point. This thread indicates that the Virtual Victim might also be found in the topics “Stats and Examples” and “Prescriptive Society”. The prevalence of terms such as “prosecutorial” and “misconduct” in the “Stats and Examples” topic documents implies that RoK is talking about legal misandry. Drilling into some of the terms that typify “Stats and Examples”, the term “%” is linked only to synonyms and currency (“income” and “GDP”) rather than stats regarding suicide rates or other purported indicators of misandry. In analysing the terms that we would expect to show links to the Virtual Victim, insights can be gained into the topic. Querying “misandry”, a link can be seen to discrimination (“bigoted”, “sex-based”, and “bias”) and decline (“wrecking”, “browbeaten”, “depreciatory”), along with the methods that are used to spread misandrist views (“grossly”, “disseminated”, “insistence”, and “pervasive”).

With paternal rights and the role of a powerful father being talking points within TRP, drilling into them confirms their importance to members. The abstract “fatherhood” is linked to neutering arguments, like “harming”, “herbivore”, “wretch”, and “detrimental” (Mountford 2018b); the ideal father figure, with terms like “testosterone-fuelled” and “dominate”; the cause of the neutering (“feminsms”, “sex-based”); and the proof of these conclusions (“eye-opener” and “studies”). Furthermore, the perpetrators of this trend, indicated by the terms “institutional”, “governmental”, and “court”, show links to heavy handedness (“top-down”, “enforced”, and “brutality”), as well as to left-wing philosophy that is driving this change (“Gramsci” and “egalitarianism”) and with descriptions of the institution enacting this discrimination (“bureaucracy”, “legislation”, “modernization”, “reform”, and “ruling”). The analysis shows that while the father is important to TRP, the current state of fatherhood is not one that is respected; it is presented as browbeaten by modern feminism via the mechanisms of legal bureaucracy.

As this analysis framed RoK as predominantly a Cyberlad masculinity, can we expect to see policing of masculinity solely in cyber masculinities? As noted earlier, the policing comes in a prescription of a traditional masculinity, with PUA modifications. Roosh’s background in PUA would explain the scale of impact of PUA, with the topic being present in 35% of documents. This proscriptive masculinity is shown in the rejection of the left and of “cuckservatism”. In TRP, this is termed “blue pill” masculinity, which would make up the othered masculinity that is presented as the alternative to taking “the Red Pill”. The data demonstrates that this policing is something inherent within all the topics that are identified as a methodology, to prove the topics importance.

Taking a selection of terms that would be expected to be typical of this policing of masculinity—“blue pill”, “loser”, “beta”, and “orbiter” (Mountford 2018b)—the LDA model can be queried for what topics are most representative within the selection. The result is that these policing terms are most strongly representative of “Personal Relationships Are Political”, with a phi similarity of over three times that of any other potential topics. Thus, it can be understood that the policing of masculinity is

done through the prescriptions on how to build relationships, and the recommended way to act in these relationships. Furthermore, the political slant of the “Personal Relationships Are Political” topic positions a generic woman and a generic man, rather than the audience. This might be understood as homosocial policing being done through the presentation of a masculine norm that is criticised, and that audiences are encouraged to deviate from it. This would resonate with the outsider positioning of the Alt-Right, increasingly rejecting a centralist or socially acceptable views (Lyons 2017).

Analysing the associations of these expected terms, some of these conclusions can be seen to resurface. “Beta” is linked with emotional dependency, with terms like “entitled” and “validation” (Mountford 2018b). “Loser” is linked with inconsistency and a lack of self-discipline, through terms like “excuse”, “whine”, “lazy”, and “bullshit”, along with the type of partner they can expect to get (“princess”, “pedestal”, “treat”, “expecting”, and “bitch”). The themes linked with the negative terms are a lack of self-discipline, treating women the wrong way (according to TRP), and women of perceived low value. These themes imply that the masculinity that is advised against will not result in achieving the aims that the audiences have (having more fulfilling relationships and performing a “better” masculinity); instead, these aims will either be subverted by audience members obtaining women of perceived low value, or by the masculinity being performed poorly or badly in some way.

Another way to answer the question of how RoK policies and advises its audience, using the data models, is to look at terms that traditionally have a negative sentiment, and understand what terms they are linked to. “Degeneracy” is linked to progressive politics and “homosexuality”, showing that hegemonic masculine homophobia as identified by Connell and Messerschmidt (2005) is still present in TRP’s neomascularity. Furthermore, the enforcement of this degeneracy is seen by RoK to be insidious and manipulative, through brainwashing (“indoctrination” and “brainwashing” (Mountford 2018b)) and subversive sneaking (“insidious”, “spreading”, and “rise”). “Bad” links to many terms likely applied to women, like “breeder” and “estrogenization”, which both dehumanise women but also reject femininity. The relational negative sentiment term “inferior” links to the biological sex argument (“biology”, “innate”, “biological”, “inherent”, and “evolutionary”), along with the results of TRP’s interpretation of innate female behaviour (“hypergammy”, “loyalty”, “selfishness”, and “denying”). From these collocations, it can be seen that TRP presents its prescriptions as the way to avoid being corrupted by insidious political movements, becoming a degenerate masculinity, or becoming a woman.

5. Conclusions

This paper proposes a novel methodology for coding and analysing large corpuses in decreased time, requiring significantly less human time invested. The results demonstrate the accuracy and insightful potential of this methodology in replicating expected findings, but also offer both novel insights and quantitative metrics than have the potential, when compared to a control corpus, to yield quantitative reliability to qualitative methodology. Furthermore, while it was not the principal aim, the potentially unpleasant nature of reading and comprehending violent or malevolent texts that Gotell and Dutton (2016) describe can be significantly negated with less in-depth engagements between the researcher and the texts.

The implications of this methodological successes are twofold. Firstly, that research of smaller and more extreme groups is feasible. Splinter groups that are likely to become crucial in later development can have their entire textual output analysed in ways that can be mapped against traditional research, because there is significantly less cost associated with open and closed coding methodology. Secondly, this methodology is able to ingest entire corpuses with negligible costs, thus removing potential sampling bias.

By using NLP models, this paper was able to effectively replicate pre-existing topics and provide deep insight. Topic modeling can demonstrate that TRP’s masculinity is based around self-development and societal changes. Furthermore, NLP can give insights into specific conceptual framings, such as the importance of historical male leaders, the perceived actions of feminism, and the

importance of PUA in achieving the goals of their audience. The data models succeed in both correctly identifying overlap and representative terms within core topics, and providing effective information retrieval when queried for simple and complex collocations.

In analysing the overall success of the data models used, it is undeniable that they flourish in the replication of proposed topics and themes, using both manual open and closed coding. The Word2vec model was perhaps the most useful tool in this research. The speed of query and insight from similar words facilitated interpretation and specific insights from the similar terms, such as TRP's presentation of how they perceive their opponents to be spreading their views (sneakily and underhandedly), and to prove the links between descriptions of women and objectivising terms. By using the Word2vec model to dig for further insight into relevant terms, as with the term "princess", the model was also used to test for expected links, as with "degenerate". The ability to replicate the Madonna/whore dichotomy and objectification re-enforces the reliability of the models proposed. With this methodology, it was possible to gain insights into the connotations and contexts of core terms, in order to re-enforce conclusions. Further work can be done with visualising this data and applying a statistical rigour to incorporate the cosine similarity, and not just similar terms.

Using the LDA topics to classify a query proved invaluable in both understanding the nuances of the topics and in understanding the query themselves. The ability to classify queries allowed for finding themes that were restricted to a topic, but which were not the most prominent strand of that topic; for example, with homosocial policing, the expected terms were within the "Personal Relationships are Political" topic, but because of the scope of this topic the terms were eclipsed without the ad-hoc query classification. For further improvement of the methods, using a graphical representation of the output of a single or multiple queries could aid the researcher in interpreting insights, while a statistical method could aid in further data mining within the returned similar terms and the topics identified. With the association, overlap, and frequency analysis of these topics, it was possible to understand the position, use, and importance given to each topic culminating in a deeper understanding of how RoK will write an article. It was most useful for understanding the co-occurrence of topics, which are common and therefore form a centre for RoK's content, and understanding which topics are often found together and which rely on each other for their arguments.

The further usage of these methodologies can extend beyond academia, to include automatic addition to extremist programmatic ad exclusion lists—such as "Excluding White Supremacist and Alt-Right Sites On Adwords" ([newwhyweb 2017](#))—early identification of users influenced by extremist media, and for auto-moderation or augmented moderation of sites based on user generated content, such as Reddit, which continues to have problems with extremist content ([Chandrasekharan et al. 2017](#)). For future research, the use of models needs to be standardised for inter-research comparability, as this is the first paper using this methodology to study reliability and the comparison of different communities and corpora.

Supplementary Materials: All code, and analysis notebooks are available online at <https://github.com/ChamRoshi/Topic-Modelling-The-Red-Pill>. The corpus and data models are available online at <https://zenodo.org/record/1168790>.

Conflicts of Interest: The author declares no conflict of interest.

References

- Beynon, John. 2001. *Masculinities and Culture*. London: McGraw-Hill Education.
- Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing With Python: Analyzing Text With The Natural Language Toolkit*. California: O'Reilly Media, Inc.
- Biricik, Alp, and Jeff Hearn. 2009. GEXcel Work in Progress Report Volume VI. Paper presented at GEXcel Theme 2: Deconstructing the Hegemony of Men and Masculinities Conference, Linköping, Sweden, April 27–29; pp. 1–233. Available online: <http://eprints.lancs.ac.uk/71687/1/GEXcel.pdf> (accessed on 3 February 2018).
- Boyd, Danah Michele. 2008. *Taken Out of Context: American Teen Sociality in Networked Publics*. Berkeley: University of California.

- Chandrasekharan, Eshwar, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined through Hate Speech. Paper presented at ACM on Human-Computer Interaction 1, no. CSCW: 31, New York, NY, USA, November 1.
- Connell, Robert W., and James W. Messerschmidt. 2005. Hegemonic masculinity: Rethinking the concept. *Gender & Society* 19: 829–59.
- Conte, Rosaria, Nigel Gilbert, Giulia Bonelli, Claudio Cioffi-Revilla, Guillaume Deffuant, Janos Kertesz, Vittorio Loreto, Helen Susannah Moat, Jean-Pierre Nadal, Angel Sanchez, and et al. 2012. Manifesto of computational social science. *European Physical Journal-Special Topics* 214: 325–46. [CrossRef]
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41: 391. [CrossRef]
- Dowsett, Gary W., Herukhuti Williams, Ana Ventuneac, and Álex Carballo-Diéguez. 2008. Taking it like a man': Masculinity and barebacking online. *Sexualities* 11: 121–41. [CrossRef]
- Forsythand, Eric N., and Craig H. Martell. 2007. Lexical and discourse analysis of online chat dialog. Paper presented at 2007 IEEE International Conference on Semantic Computing (ICSC 2007), Irvine, CA, USA, September 17–19; pp. 19–26.
- Ging, Debbie. 2017. Alphas, Betas, and Incels: Theorizing the masculinities of the Manosphere. *Men and Masculinities*. [CrossRef]
- Gotell, Lise, and Emily Dutton. 2016. Sexual violence in the 'manosphere': Antifeminist men's rights discourses on rape. *International Journal for Crime, Justice and Social Democracy* 5: 65–80. [CrossRef]
- Hendriks, Eric C. 2012. Ascetic Hedonism: Self and sexual conquest in the seduction community. *Cultural Analysis* 11: 1–16.
- Jacobi, Carina, Wouter van Atteveldt, and Kasper Welbers. 2016. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism* 4: 89–106. [CrossRef]
- Jane, Emma A. 2017. Systemic misogyny exposed: Translating Rapeglish from the Manosphere with a Random Rape Threat Generator. *International Journal of Cultural Studies*. [CrossRef]
- Johnson, Deborah G. 1997. Ethics online. *Communications of the ACM* 40: 60–65. [CrossRef]
- Kelly, Annie. 2017. The Alt-Right: Reactionary rehabilitation for white masculinity. *Soundings* 66: 68–78. [CrossRef]
- Kendall, Lori. 2002. *Hanging Out in the Virtual Pub: Masculinities and Relationships Online*. Berkeley: University of California Press.
- Khandelwal, Shashikant, and Susumu Harada. 2004. *Automatic Topic Extraction and Classification of Usenet Threads*. Stanford: Stanford University, Available online: http://nlp.stanford.edu/courses/cs224n/2003/fp/kshashi/project_khandelwal_harada.doc (accessed on 1 November 2017).
- Leonard, Peter. 2014. Mining Large Datasets for the Humanities. Paper presented at IFLA WLIC 2014-Lyon-Libraries, Citizens, Societies: Confluence for Knowledge, Lyon, France, August 16–22.
- Light, Ben. 2013. Networked Masculinities and social networking sites: A call for the analysis of men and contemporary digital media. *Masculinities & Social Change* 2: 245–65.
- Luthra, G. S. 2017. Why Is Donald Trump Doing Israel's Bidding?—Return Of Kings. Returnofkings.Com. Available online: <http://www.returnofkings.com/127418/why-is-donald-trump-doing-israels-bidding> (accessed on 12 December 2017).
- Lyons, Matthew N. 2017. *Ctrl-Alt-Delete: The Origins and Ideology of the Alternative Right*. Somerville: Political Research Associates.
- Messner, Michael A. 2016. Forks in the Road of Men's Gender Politics: Men's Rights vs Feminist Allies. *International Journal for Crime, Justice and Social Democracy* 5: 6–20. [CrossRef]
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv*. preprint arXiv:1301.3781. Available online: <https://arxiv.org/abs/1301.3781> (accessed on 18 November 2017).
- Mountford, J. B. 2015. Creating Masculinities Online: Bronies and the Red Pill. Master's dissertation, University of Sussex, Brighton, UK.
- Mountford, J. B. 2018a. Topic Modeling the Red Pill—Datasets. *Zenodo*. [CrossRef]
- Mountford, J. B. 2018b. Topic Modeling the Red Pill—Analysis. *GitHub Repository*. Available online: <https://github.com/ChamRoshi/Topic-Modelling-The-Red-Pill> (accessed on 3 February 2018).

- Mowlabocus, Sharif. 2006. Chat Rooms and Back Rooms: Same-Sex Desire, Cyberspace, and Contemporary British Gay Male Sub-Culture. Ph.D. thesis, University of Sussex, Brighton, UK.
- newwhyweb. 2017. Excluding White Supremacist & Alt-Right Sites on Adwords. *New Why*. Available online: <http://www.newwhyweb.com/white-supremacist-alt-right-adwords-exclusion/> (accessed on 20 October 2017).
- Price, Rob. 2015. The Web's Most Infamous Misogynist Regrets Nothing. *The Daily Dot*. Available online: <https://www.dailydot.com/irl/roosh-v-mens-rights-sexism-misogyny/> (accessed on 4 January 2018).
- Rehurek, Radim, and Petr Sojka. 2010. Software framework for topic modelling with large corpora. Paper presented at the LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta, May 22.
- Richardson, Leonard. 2013. Beautiful Soup. Crummy.com. Available online: <https://www.crummy.com/software/BeautifulSoup/> (accessed on 11 September 2017).
- Sardar, Ziauddin. 1995. alt. civilizations.faq cyberspace as the darker side of the West. *Futures* 27: 777–94.
- Schmitz, Rachel M., and Emily Kazyak. 2016. Masculinities in Cyberspace: An Analysis of Portrayals of Manhood in Men's Rights Activist Websites. *Social Sciences* 5: 18. [CrossRef]
- Southern Poverty Law Center. 2012. Misogyny: The Sites. In *Intelligence Report* 145. Montgomery: Southern Poverty Law Center.
- Tirunillai, Seshadri, and Gerard J. Tellis. 2014. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research* 51: 463–79. [CrossRef]
- Valizadeh, Daryush. 2015. How to Stop Rape. RooshV.Com. Available online: <http://www.rooshv.com/how-to-stop-rape> (accessed on 12 December 2017).
- Van Dijck, José. 2009. Users like you? Theorizing agency in user-generated content. *Media, Culture & Society* 31: 41–58.
- Wu, Xindong, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. 2014. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering* 26: 97–107.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).