Discourses of the Fringe Right:

Analysing Discourse and [De]legitimation on Gab.ai

Alex Phillip Lyng Christiansen

Aalborg University

Author Note

Alex P. L. Christiansen, School of Culture and Global Studies, Aalborg University

Correspondence concerning this assignment should be addressed to Alex Christiansen, 6 Mill View House, LA1 1AU Lancaster, Lancashire

Contact: aplc13@student.aau.dk

Supervised by Christina Schoux Casey

My most sincere appreciation is owed to the following…

To Christina, whose optimism and helpfulness has meant more than she knows, in the face of a difficult and ever-changing task. To Edward Dearden and Jeremy Blackburn, without whom this paper would be weirdly lacking in data. To my one and only Karolina, whose attitude towards my never-ending cycle of work has been remarkably patient and supporting.

And finally, to Peter and Emma Thomsen, whose generous and continuous support has been invaluable and has indirectly led to many of the best things in my life. Thanks for taking a chance on me.

Abstract

Given the worrying growth of anti-Semitic and racist tendencies and actions in both Europe (Ward, 2018) and the US (Anti-Defamation League, 2018), it is important to continuously evaluate the developing communities on social media, where a majority of recruitment takes place. In spite of its massive growth, however, there has until now been very few attempts at analysing the discourse community present on the 'alt-right' twitter alternative, Gab.ai. As a tentative step towards rectifying this, the present study analyses the most liked, disliked and controversial sentiments on Gab. As such, using a sampled corpus which spans Gabs existence up until 2018, the study presents a contextually sound image of the discourses which are legitimised and de-legitimised on the website.

The results of the study support allegations and previous observations of alt-right activity, highlighting recurrent use of extremist right-wing and nationalist terms, such as 'deep state', 'fake news' and 'white genocide'. Additionally, the conversation around 'free speech' shows a community which supports unlimited free speech, up to and including 'hate speech' and reject the notion of 'hate crime' and 'hate speech' as valid terminology entirely.

The discursive make-up of the three analysed sub-corpora also suggests a systematic use of downvoting and negative discursive evaluation to disparage opposing, i.e. left-leaning, political views and discourses in general. The disliked sentiments serve as negative mirror of those which are deemed legitimate, including people who defend the right of leftist organisations like antifa and those who criticise conspiracies or white nationalists in general.

The study concludes by proposing that the discourses found on Gab reflects a community which actively delegitimises opinions opposite to those of the majority, creating an effectively homogenous community. This observation ultimately suggests that further studies of Gab as a corpus of primarily alt-right interactions are valid, given that most occurrences of counter-discourse may be successfully separated using the metrics presented in the study.

Keywords: Discourse Analysis, Corpus Linguistics, Legitimation, Gab

Table of Contents

Abstract iii
1. Introduction 1
2. Context 3
2.1 From white nationalism to the alt-lite 3
2.2 The ‘free speech’ Network 5
2.2.1 Previous Studies 6
3. Theory and Methodology 7
3.1 Corpus Assisted Discourse Studies 8
3.2 Critical Discourse and Legitimation 8
3.3 Previous works 9
4. Data 10
4.1 Data Collection 10
4.1.1 Ethics 12
4.2 Sub-Corpora 12
4.3 Reference Corpora 13
4.4 Stop Words 14
5. Methods 14
5.1 Software 14
5.2 Concordance 15
5.3 Keywords 15
5.4 Collocation 16
5.5 Clusters 18
6. Analysis 18
6.1 [De]legitimation 18
6.1.1 *Like* 19
6.1.2 *Dislike* 31
6.2 ‘|x| She’ 36
6.2.1 *Controversial* 37
7. Discussion 39
7.1 Discourse and Legitimation 40
7.2 A Corpus of the Alt-Right 42

8. Conclusion 42
9. Issues and Future Studies 43
9.1 A note on Bots 43
9.2 A Note on Multimodality 44
Works Cited 46

Table of Figures and Tables

Figure 1. Screenshot of Gab.ai homepage 5
Figure 2. Example of collocate network of ‘example’ in #LancsBox 17
Table 1. Gab API 11
Table 2. N-gram Keywords in *Like* ranked by freq. 19
Table 3. Collocates of ‘free speech’ in *Like* ranked by MI. 21
Table 4. Collocates of 'hate speech' in *Like* ranked by MI 21
Table 5. Concordances of ‘free speech’ collocating with ‘is’ in *Like* 22
Table 6. Concordances of ‘hate speech’ collocating with ‘is’ in *Like* 23
Table 7. Concordances of 'hate speech' collocating with 'is' in *Like* 25
Table 8. Collocates of 'fake news' ranked by MI 26
Table 9. Concordances of 'fake news' collocating with 'is' in *Like* 27
Table 10. Concordances of 'western civilization' collocating with 'is' in *Like* 28
Table 11. Concordances of 'white genocide’ and ‘civil war' in *Like* 29
Table 12. Concordances of 'red pill' in *Like* 30
Table 13. Concordances of 'deep state' in *Like* 30
Table 14. N-gram Keywords in Dislike ranked by freq. 31
Table 15. Collocates of 'free speech' in *Dislike* ranked by MI 33
Table 16. Concordances of 'free speech' collocating with 'is' in *Dislike* 34
Table 17. Concordances of 'hate speech' collocating with 'is' in *Dislike* 34
Table 18. Concordances of 'fake news' in *Dislike* 35
Table 19. Collocates of 'she' in *Like* and *Dislike* 36
Table 20. Concordances of 'she' in *Dislike* 37
Table 21. Concordances extracted from *Controversial* 39

1. Introduction

Until recently, alt-right and alt-lite communities have generally been allowed to continue operating within the public spheres, to some degree avoiding the scrutiny of 'mainstream' social media, such as Facebook, Twitter and YouTube, but this would seem to be coming to an end. While some key individuals of the alt-right, such as Richard Spencer, the creator of the term 'alt-right', have been left alone by Twitter (although removed from Facebook), others have not been so lucky. Consequently, large alt-right and alt-lite communities such as The Proud Boys in the US and The English Defence League in the UK have been forced to relocate.

Given a long-held belief by researchers in the field (e.g. Goodwin, 2011) that the far right is connecting with 'a new generation', observing these developments is paramount to the continued appliance of preventative measures. Earlier observations assessing the threat of white nationalism concluded that one of the primary inhibitors of growth in the movement has been the failure to advance to a phase of 'post-awakening' (Berger, 2016, p. 25) in which members are further radicalised on a wider scale. The reason, Berger surmises, is that white nationalism and far right activity is too fragmented and lacks a singular communal space. Two years later, such a space now arguably exists, and following the tragic attack on a synagogue in Pittsburgh on October 24th, in which 11 people lost their lives and 6 were injured, media attention quickly turned to the place where the alleged gunman, Robert Bowers, had made clear allusions to violent action only hours before the attack began: Gab.ai.

Touting itself a 'defender of free speech', Gab has quickly attracted the attention of people who have been rejected by Facebook and Twitter, many of which are members of various parts of the alt-right and alt-lite (see e.g. Yates, 2016; Lee, 2017; Brandom, 2018). although the creator of the site, Andrew Torba, has often claimed that Gab was never meant to be a place specifically for people of this persuasion, the growing popularity amongst banned Twitter-users has led many outlets to name Gab the 'alt-right social network' (Bray 2016; Hess 2016; Keyser 2018), citing the violent and often hateful speech which has seen Gab removed from many hosting platforms, including Google's own (Toor 2017; Brandom 2018). It has also been broadly theorised that Gab's logo, a green frog, serves as a direct reference to 'pepe the frog', a meme which was heavily co-opted by alt-right and white nationalist communities throughout the election period of 2016 (ADL, "Pepe the Frog").

Given the worrying growth extremist right, anti-Semitic and racist tendencies and actions in both Europe (Ward, 2018) and the US (Anti-Defamation League, 2018), it is

important to continuously evaluate the developing communities on social media, where a majority of white nationalist recruitment takes place. As Twitter and Facebook are cracking down, Gab functions as a new possible site of analysis. Furthermore, given the purported homogeneity of Gab, the website may facilitate a corpus of far-right discourse, presumably without having to filter semantic differences or sarcastic and ironic ridicule of the discourse from opposing groups, which previous studies have found to be prevalent on Twitter (Christiansen, unpublished). Such a corpus could, in turn, facilitate more direct access to the day-to-day interactions that legitimise and normalise hateful populist rhetoric and further extremist actions. It is to this end that the present paper attempts a broad-scale examination of discourse present on the so-called 'Alt-right Twitter' and 'white nationalist filter-bubble', Gab.ai.

In the two years since its inception, Gab has grown dramatically, projecting a user-base of 800.000 by the end of 2018. But despite its massive growth in users and increasing importance to the mainstream, there has been only two scholarly examination of Gab.ai; Zannettou et al (2018) and Lima et al. (2018), both of which fail to provide much beyond a purely statistical image of the site and its relation to hateful content. In an effort to provide further characterisation, then, the present study examines the discourse of Gab at a more granular level, focusing specifically on identifying the active (de)legitimation on opinions which reflect the community purported to exist almost exclusively on Gab. Given the website's purported goal of supporting the nebulous concept of 'free speech', the study also attempts to ascertain the way the community legitimises particular views of 'free speech' and 'hate speech'. Overall, then, the examination combines elements of corpus linguistics and discourse analysis to analyse the topics, behaviour and attitudinal values which are codified and enacted on Gab, in order to answer the following questions:

1. Which discourses are typified and legitimised, as opposed to rejected or delegitimised, by the Gab community, and in what ways do the findings support or reject previous observations and accusations of alt-right homogeneity on the site?

2. What obstacles, if any, hinder the utilisation of Gab as the foundation for a corpus of alt-right social media interactions?

The study is split into 8 parts. This first section has presented the hole in the literature on the social media site Gab, in the face of its continued growth and importance to observing fringe right communities. Section 2 adds further context to focus of the study by examining the differences between white nationalism, the alt-right and the alt-lite and presents a brief introduction to the history and previous studies of Gab.ai. Section 3, 4 and 5 details the make-up of

the study, with section 3 introducing the methodology, section 4 introducing the data and data-collection approach and section 5 detailing the various methods from corpus linguistics utilised in the analysis. Section 6 presents the analysis and is split into three sub-sections, each of which deals with the analysis of a particular sub-corpus. Following this, section 7 discusses the results of the analysis, particularly in relation to the previous studies detailed earlier in the paper, directly followed by a section 8, which summarises the findings made in the analysis and offers some concluding remarks. Finally, the study ends with a brief discussion of issues and future studies in section 9.

2. Context

The following section serves to provide context to the topic of analysis. Here, I present a brief introduction to the concepts of 'alt-right' and 'alt-lite' and an introduction to the website http://www.gab.ai, including an overview of previous works analysing the site.

2.1 From white nationalism to the alt-lite

To be clear, while the present study leans on characterisations of the political right as belonging primarily to either the alt-lite, alt-right or, in rare instances, self-professed white nationalist and neo-Nazi groups, these definitions are fluid and changing in a manner that defies a constant. This fluidity extends beyond simply the definition of the groups as well, with many members of either group transferring or straddling the lines between groups at various points in time. As an example, former Breitbart tech editor and alt-right spokesperson Milo Yiannopoulos has since expressed issues with the group's antisemitic leanings and is now commonly considered part of the alt-lite. The term 'alt-lite' was primarily coined in opposition to the extreme nature of the alt-right, an opposition often attributed to the more outright connections a prominent spokesman for the alt-right, Richard Spencer, made to the symbolism and honorifics of Nazi ideology (Dearden 2016).

Although the present study does not concern itself with white nationalists or the KKK per se, it is an unavoidable truth that these more 'old-school' fascist groups have achieved some form of mainstream appeal through continued proliferation of memetic content and the rise of the so-called 'alt-right' (see e.g. Berger, 2016, Gray, 2018; Lima et al., 2018). Indeed, Berger (2016, pp. 25-26) even notes a growing overlap between online culture and white nationalism, remarking that a still increasing 'trolling culture' is "amplifying the presence of white nationalist content on the social media" creating further confusion as to whether these

users are committed white nationalists or just trolls doing it – in the parlance of the community – 'for the lulz'.

Broadly speaking, however, examinations of white nationalist extremism have so far failed to recognise the importance of a rapidly growing non-extreme sphere, now commonly referred to as the 'alt-lite', in the continued legitimation of ethnonationalist values. While the alt-lite is mostly associated with memetic ridicule and a general distaste for identity politics and political correctness, they are also at times a source of harassment campaigns, disinformation, and civic nationalism, and tinged with a hunger for conspiracy theories. Going further, the alt-lite remains the most obvious and prime source of recruitment for the hardcore alt-right, as noted by prominent white nationalist voice Greg Johnson, who stated that the alt-lite should be encouraged as "most" will be "White Nationalist material" (Johnson, 2017).

The term alt-lite, sometimes known as the *alt-light* or the *new right* contains the broadest orbit of what is commonly referred to as the 'alt-right' and is often defined in opposition to it. Whereas the alt-right self-proclaims as racists (or 'race realists') and argue for the inevitable necessity of an ethnic cleansing in the face of a 'white genocide', the alt-lite admits no such allegiance and prefers to frame immigration within a discourse of economic impact and civic nationalism. The alt-lite openly rejects concepts such as globalisation, open borders, feminism, and political correctness, and hails masculinity, IQ, and 'free speech' amongst other things (ADL "From Alt Right to Alt Lite: Naming the Hate").

At first, separating the alt-right from the alt-lite may appear pointless, but it is important to keep in mind both that the alt-lite is *not* the alt-right and that, paradoxically, they may become so at any moment. As already alluded to, the alt-lite is considered less extreme, preferring memetic ridicule, harassment campaigns and general misogyny over direct violent attacks such as the ones carried out by white nationalists in Charlottesville in 2017 and Pittsburgh in 2018. At the same time, it is far larger, with prominent alt-lite symbols such as Milo Yiannopoulos and Paul Joseph Watson boasting followers in the millions, reaching audiences far beyond those of self-professed white nationalists. It is thus through this intrinsic allure of the alt-lite, one which focuses on youthful resistance, that extremist right-wing movements may start to move towards the kind of 'post-awakening' theorised by Berger (2016). By preying on an angry and rebellious youth culture, fed up with disinformation and 'fake news', white nationalists hope to be the last stop on a longer developmental journey. As prominent white nationalist Greg Johnson (2016) writes, "if we can't convince them to come the rest of the way with this sort of setup, we are doing something wrong."

2.2 The ‘free speech’ Network

As the promise of Web 2.0 democracy has waned and hate speech has become common-place on the internet, social media platforms have been forced to reconsider their stances on moderation in an increasingly hostile public environment. This growing opposition towards harassment resulted in the de-platforming of many neo-conservative voices, resulting in mass-migration amongst those who were no longer welcome on mainstream social media and kick-starting the search for a more ‘accepting’ communal space. The answer to this search, as it turned out, would come from Silicon Valley itself, with the initial creation of Gab.ai in 2016 (Kantrowitz 2016).

According to site creator Andrew Torba, the creation of Gab.ai should be seen as a direct answer to the mass-censorship of conservative voices on social media platforms such as Facebook and Twitter. As a website, Gab bears a striking resemblance to Twitter in particular, from its horizontal Home | Notifications | Messages | Search | User bar to the timeline stretching out below the dedicated ‘new post’ window. Gab even features a similarly character-limited posting system, allowing its users 300 characters to Twitter’s current maximum of 280.

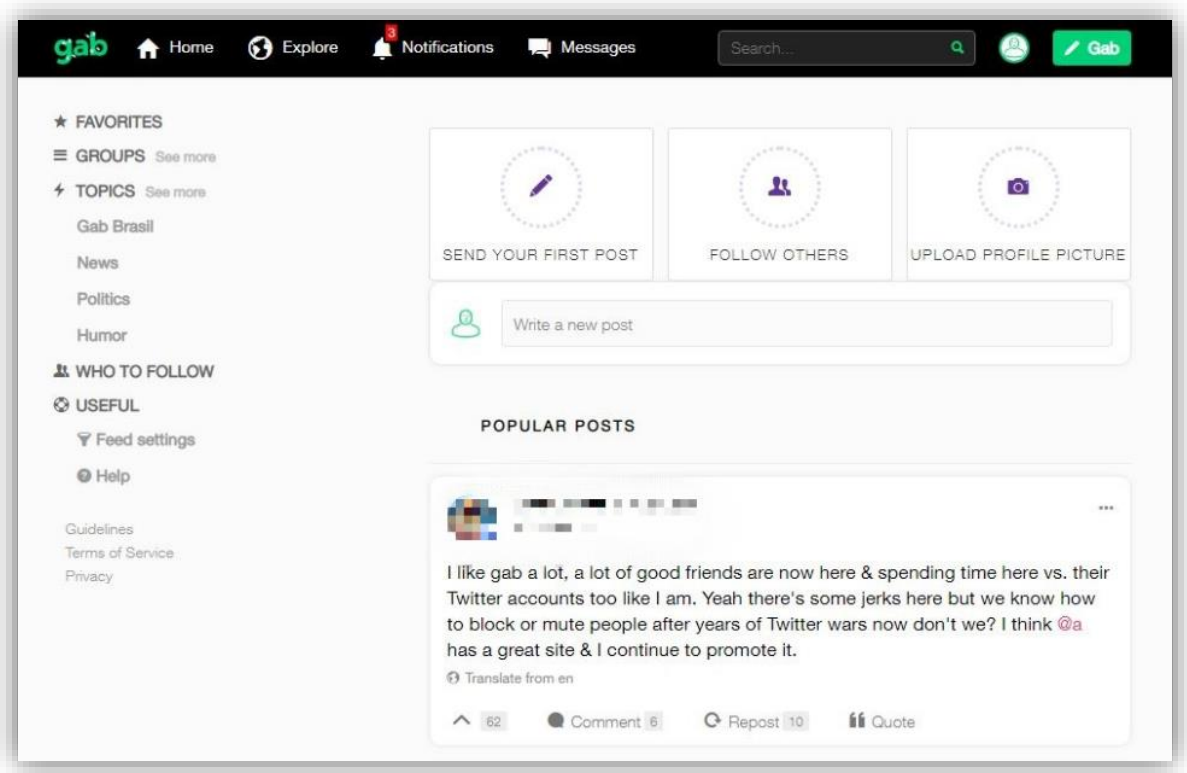

Figure 1. Screenshot of Gab.ai homepage

The true difference between Gab and its stated competition, then, is its user-base. It is hardly unfair to say that the ones Twitter rejects Gab welcomes. This, along with the site's design, leads many to define the site as a right-wing alternative to Twitter (Ellis 2016; Keyser 2018). While Gab is not alone in providing a homestead to people banned by Twitter and Facebook, the free-speech twitter alternative has proven itself different from other alt-right-positive websites such as WrongThink and ReaganBook by its continued growth, going from just 400.000 users in February 2018 to more than 700.000 as of October 2018 ("The Home of Free Speech Online").

Although Torba has consistently claimed that Gab was never meant to be a place specifically for the alt-right, previous studies have concluded that the website serves mainly conservative Caucasian men (Lima et al. 2018, p. 522) and that the site exists in a space somewhere between moderate platforms like twitter and facebook and alt-right platforms like /pol/ on 4chan and r/TheDonald on Reddit.. To this day, Gab continues to struggle with retaining the website, having previously been blacklisted by cloud hosting services such as Google and Microsoft's Azure (Brandom 2018), owing to its extremist user-base. Following continued scrutinization in the wake of the Pittsburgh attack, the website was shut down for almost two weeks in November 2018, while Gab worked with the FBI by providing information on the suspect, Robert Bower. This marked the second time that Gab was forced to officially and publicly disavow one of their users for going beyond what the website willingly defines as 'free speech'.

2.2.1 Previous Studies

Although the present paper is the first study analysing Gab from a primarily discursive point of view, previous studies have produced characterisations of various aspects of the website's community in Zannettou et al. (2018) and Lima et al. (2018). While rendering a full critique of these studies, such as they are, is outside of the scope of the present paper, I would be remiss not to consider their contributions and findings in relation to the approach taken here.

Lima et al (2018) arguably lays the foundation for characterising Gab, as it attempts to characterise both the users of the site and the content shared amongst these users. To do so, the study employs a sampled corpus of approximately 12.8 million posts and visualise several different metrics, including political leanings, racial features and extremist/hate speech tendencies, primarily by accessing the top-level information detailed in Section 4.1, Table 1. In comparison, Zannettou et al. (2018) goes slightly more in-depth in its analysis of hate

speech on Gab, by including surface-level lexical analysis of keywords and bi-grams in their significantly larger corpus of about 23 million words.

Overall, the findings in both studies lend support to the previously made assertions that Gab is a conservative echo chamber. First, using time-mapping, Zannettou et al. (2018, p. 1014) shows that Gab is a highly political network, with strong reactions to current events, especially those focused around white nationalism and support of Donald Trump. These observations are supported by the ones made in Lima et al. (2018, p. 521), which adds that users are often found to be on the conservative spectrum on a classic liberal / conservative ideological model. As both approaches rely almost entirely on single-word keywords and statistical evidence, however, discounting a few anecdotal examples in Lima et al. (p. 520), both conclusions are arguably flawed. While the results in Zannettou et al. (2018) are viable in that they highlight political activity, previous findings on Twitter in Christiansen (Unpublished) showed frequent use of counter-discourse and sarcastic discourse within communities opposed to these particular communities. Given that such use tend to use similar or even purposely identical keywords and clusters, such occurrences would theoretically fly under the radar of the approach detailed in Zannettou et al. Interestingly, in terms of political bias, Lima also finds that almost half (45.5%) are found to be moderate (Lima et al., 2018, p. 518), indicating an overall more centrist community with a conservative slant, rather than 'white genocide' conspiracy theorists concluded to be prevalent in Zannettou et al. (2018) raising questions as to which study is correct. Using facial recognition software, Lima et al (2018, p. 517) also proposes that the userbase of Gab are primarily male (67.2%) and white (76.1%), with 15.8% being Asian and 8.2% being black. While I cannot say for certain that these results are to be disputed, preliminary results from the as-yet-unreleased Twitter Internet Research Agency corpus (Twitter About, 2018) may indicate that this is the case. So far, ongoing studies show that many of the scam accounts run by the Russian agency were pretending to be Black Lives Matter or other minority-community groups. Given that these attempts appear to be made in order to further antagonise both left and right-leaning communities online, in order to sow political dissent, it is likely that this extends to Gab as well.

3. Theory and Methodology

The following section details the methodological and theoretical foundation of the paper. I begin with an introduction of the core of corpus linguistics and discourse analysis under the common umbrella of corpus assisted discourse studies, followed by a brief presentation of

discourse as it relates to legitimation. I end the section with a short comparison of previous works in the literature on social media which use similar methodology.

3.1 Corpus Assisted Discourse Studies

The methodological core of the present paper is the marriage between qualitatively-informed discourse analysis and data-driven linguistic analysis using a large collection of texts, commonly referred to as a corpus. While the qualitative part consists of arguably subjectively informed analysis of discourse, it is through the latter part that the study attempts to generalise the findings towards a broader understanding of Gab as a whole. As an approach, this reflects an epistemological belief in the combination of theory and method, corpus linguistics (Henceforth CL) and discourse analysis (Henceforth DA), and it is with this combinatory definition in mind that the paper adheres to the umbrella term 'corpus assisted discourse studies' (Henceforth CADS).

To say that something is corpus-assisted, or corpus-based, is to clarify that the approach relies on computer-assisted methods and quantitative, data-driven analytical tools, in order to explore large, often qualitatively insurmountable collections of texts. In relation to DA, corpus assisted refers to corpora being used in conjunction with a prevalent theory of discourse and is thus adherent to McEnery and McGlashan's (2013, p. 5) broader definition as "any approach to language that uses corpus data and [corpus] methods". An in-depth explanation of these methods can be found later on, in section 5.

While the application of corpus methods serves as a way to deal with the enormity of the Gab corpus, it is only through the qualitative analysis text in context, i.e. concordances, that one may render any assumptions about the meaning of the numbers presented. Although adopting a corpus linguistic approach is commonly seen as a way to reduce the inherent bias often present in ideologically motivated work looking at language, it is important to note that this does not render CL free of bias or subjectivity in general. As noted by Biber (Biber, Conrad and Reppen 1998: 4), "functional (qualitative) interpretation is also an essential step in any corpus-based analysis", and nowhere is this more prudent than in the area of discourse.

3.2 Critical Discourse and Legitimation

What started out as essentially being content analysis has slowly transformed into a more critical linguistic perspective on an online community, arguably taking on the mantle of Critical Discourse Analysis (CDA) over the less ideologically informed DA. I accept this, in the sense that CDA, as noted in Fairclough (2012, p. 12; Fairclough, 2017), can be broadly defined as a

"loosely interconnected set of different approaches" pertaining to specifically critical forms of social analysis. While the study does not purport to desire material change in a political sense, and thus fails to adhere to CDA as dialectical reasoning (i.e. Fairclough, 2017), it still arguably follows loosely in the footsteps of Fairclough's critical methodology as outlined in Fairclough (2015, p. 4) in examining a community for so-called 'social wrongs'. Rather than suggest transformative action, then, the present study is an exercise in observing social practices from a somewhat neutral stand-point, in order to confirm or invalidate previous examinations of Gab.

As such, although the present study does not include any direct form of social practice analysis or legitimation analysis, the approach to critical discourse is still largely grounded in observations of legitimation and social practice (Van Leeuwen, 2008). Since Gab has been accused of becoming a 'filter-bubble' for the alt-right, the study seeks to ascertain the validity of such a claim by testing whether the site is indeed homogenous, or whether it allows a plurality of discourses. It is to the broadest degree, then, that the paper can be said to analyse social practices; in looking for hidden 'regulation' of discourses reflected in the social practice of downvoting and upvoting, rather than analysing particular behaviour amongst individual users on the site. The paper is, in this sense, attempting to critically evaluate the use of downvoting and upvoting as a tactical tool by a conscious, possibly homogenous community.

3.3 Previous works

While there are numerous examples of previous work analysing hateful communities, hate speech and social practice on social media through the lens of corpus linguistics, (e.g. Hardaker, 2010; Hardaker and McGlashan, 2016; McEnery, McGlashan and Love, 2015; Chetty and Alathur, 2018) and even more simply utilising social media for corpus analysis, these examinations have, for the most part, been limited to smaller case studies of topics or events. Additionally, though numerous studies use big-data approaches to analysing Twitter, Facebook and even Gab, as exemplified in Lima et al. and Zannettou et al., none of these studies fail to apply a granular discourse-focused perspective. In contrast to previous work, then, the present study aims to provide a semi-comprehensive overview of the most prevalent discourses across an entire website. As covered in the data-overview found in the following section, this leads to the employment of a far larger corpus than can be found in any of the above-mentioned discourse-focused examples, consequently requiring a far more flexible set of tools than what is commonly seen in linguistic analyses.

4. Data

The following section details the data utilised throughout the present study, as well as the procedure of collecting the data itself and splitting it into the three utilised sub-corpora. It also briefly touches on the ethics of using natural language data from a social media site like Gab.

Only a small part of the overall data was utilised in the analysis, as detailed in the sub-corpora breakdown found in section 4.2, making up just 61,000 posts and just 498,000 words. The full data available to the present study includes just over 23 million rows (23,252,000), with each row representing a post made by a user on the site. As detailed in section 4.1, the large majority of these, about 22,848,000 rows, were taken from an existing corpus, while a small additional amount, about 404,000 rows, was extracted as part of the present study. A de-duplication search was done to remove all identical posts at the 'body' level (see explanation below), effectively removing almost 8 million rows of repeat data, showing the amount of repeat data in the overall corpus. Additionally, an unknown number of these words are still likely to be links or hashtags.

4.1 Data Collection

The data collection procedure has been somewhat problematic. The reason it has been successful, then, is in part due to the tireless help of Ed Dearden, as well as the generous contribution of external data from Zannettou et al (2018), courtesy of Jeremy Blackburn. Since no prior study has given a detailed rundown of the information contained in the API, this section presents a rudimentary rundown of that data, in a manner similar to the Twitter-rundown provided in Hardaker and McGlashan (2016).

To begin with, it should be stated that Gab.ai has an open API, allowing anyone to scrape any public posts – collegially referred to as 'Gabs' – made on the site using a simple posting ID. As a random example, *https://gab.ai/posts/500060* gives us post number 500,060. The information included is fairly comprehensive, as well, allowing for far more detailed analyses than what is afforded by Twitter's limited data-hose approach which is detailed in the Twitter Developer API (Twitter, 2017). Sadly, the freedom afforded by the Gab API is somewhat occluded by the fact that no documentation is available for the way it is constructed, leading researchers to largely fumble blindly in an effort to find the best approach for extracting large amounts of data. The resulting information can be seen below, in table 1.

Data	Value	Description

id	number	Shows post ID, the first post would be '1'
created_at	**date**	**Shows the post creation date, e.g. 2016-09-26**
revised_at	date	Shows date(s) of revision, e.g. 2016-09-26 - depends on "edited"
edited	true/false	Shows whether any edits have been made
body	**text**	**Shows the part of the post consisting of written text**
only_emoji	true/false	Shows whether the post consists solely of emoji
liked	true/false	Shows whether the post has been liked
disliked	true/false	Shows whether the post has been disliked
bookmarked	true/false	Shows whether the post has been bookmarked
repost	true/false	Shows whether the post has been reposted
reported	true/false	Shows whether the ported has been reported
score	number	Shows the score of the post, measured by calculating "like"-"dislike" counts
like_count	**number**	**Shows the amount of times the post has been liked**
dislike_count	**number**	**Shows the amount of times the post has been disliked**
reply_count	number	Shows the amount of replies made to the post
repost_count	number	Shows the amount of times the post has been reposted
is_quote	true/false	Shows whether the post is a quote
is_reply	**true/false**	**Shows whether the post is a reply**
is_replies_disabled	true/false	Shows whether the poster has disabled replies
embed	all*	Shows any embedded posts, including all previously noted information making up a post
attachment	image/video	Shows the information related to any images, videos or gifs included in the post
category	number	Shows whether a post is part of a 'category'
category_details	**topic**	**Shows the topics/categories with which the poster has associated the post, e.g. 'News', 'Religion' - depends on "category"**
language	language	Shows the language associated with the poster
nsfw	true/false	Shows whether the post is considered 'safe for work', indicating either porn or extreme violence
hidden	true/false	Shows whether the post has been 'hidden' by the poster, disallowing people from seeing it
is_premium	true/false	Shows whether the poster is a 'premium user'
is_locked	true/false	Shows whether the post has been locked, barring any further replies
user	number	Shows user ID, the first user would be '1'
name	name	Shows the public name of the poster
username	**name**	**Shows the login name of the poster**
verified	true/false	Shows whether the poster is 'verified' as a public person
is_donor	true/false	Shows whether the poster is a donor from previous donation campaigns
is_investor	true/false	Shows whether the poster is an investor from previous investment campaigns
is_private	true/false	Shows whether the post has been set to 'private', barring any but the poster's accepted followers from seeing their posts.
replies	all*	Shows any replies made to the post, including all previous noted information making up a post

Table 1. Gab API

From this list of data, the study relies on only 7 of the values found in each post, all of which are indicated with bold in Table 1: *Creation date* allows us to create a timeline of changes in

discourse over time. *Body* gives the text-portion of the post and is a necessary requirement for the purpose of any text analysis. *Like*/*dislike count* shows us what types of posts are popular on the site and which ones are not, allowing us, amongst other things, to analyse the type of discourses which are foregrounded and backgrounded through (de)legitimation tactics by the community. *Reply* allows us to filter for posts which are original, rather than replies – this is necessary, as the body of reply posts often lack the context afforded in their original html format, where all of the original post will also be embedded. *Category* allows us to filter for posts belonging to a particular topic, such as *news* or *politics*. Finally, *username* allows us to find all posts from a single user by filtering according to name.

4.1.1 Ethics

When working with data obtained from natural interactions, whether online or offline, permission is needed in cases where the information may be re-distributed or published. As noted in the 'recommendations on good practice' in the BAAL (2006, p. 5) guidelines, however, "observations in public places is a particularly problematic issue" and "In the case of an open-access site, where contributions are publicly archived, and informants might reasonably be expected to regard their contributions as public, individual consent may not be required" Since it is clearly impossible to obtain permission from every individual included in the written corpus collected from Gab, I resort to the next best thing and anonymise the data. Even in examples directly taken from the corpus, where a post is relayed in its entirety, no username, user-ID or links which identify the profile have been included in the study. If any individual included in the data object to their inclusion, this inclusion will be negotiated and possibly removed

4.2 Sub-Corpora

A sub-corpus is a corpus which has been built by taking pieces of an overarching main corpus. The main corpus here consists of all collected Gab posts. Sub-corpora are often used to highlight more specific terms, topics or communities found within a corpus, by isolating aspects which may relate to these. Consequently, they often consist of a far smaller portion of the overall corpus. Given that the present study presents a far more in-depth view of the discourse present on Gab, doing analysis on the entire corpus would be methodologically challenging without access to far more resources. Instead, then, the study takes advantage of the 'like' versus 'dislike' feature on Gab to create three separate sub-corpora for analysis: *Like*,

Dislike, and *Controversial*. Throughout the paper, these terms will appear italicised whenever they reference their respective corpora.

To begin with, the *Like* corpus is fairly heavily narrowed to only posts which have an overall score of 150 or more upvotes, leaving only the most prevalently agreed-with posts. This cut-off point is mostly down to the fact that the overwhelming amount of posts do not receive dislikes, and approximately one in ten posts received more than 100 likes. Broadening the data would be trivial but processing it would require a far more time and a better dedicated program for analysis. The second corpus, *Dislike*, includes all the posts which have both 30 or more dislikes *and* has an overall negative score of at least 30 when subtracting likes from dislikes. Lastly, the *Controversial* corpus includes any post with 30 or more dislikes but an overall *positive* score, indicating a split in the opinion of the community, although sometimes overwhelmingly to one side or the other. It should be noted that while the cut-off points are essentially arbitrary, the choice to enforce fairly strict measures means that any observation found in the employed sub-corpora will have been actively deemed either legitimate or illegitimate by at least 30 people. As such, what each sub-corpus reflects is not only language which can be found on Gab, but language which has been evaluated by at least some portion of the community as either legitimate or illegitimate.

4.3 Reference Corpora

When analysing keywords, as is done numerous times throughout the paper, a key requirement is the employment of a reference corpus. While the choice of reference corpora can differ significantly based on the goal of analysis, the utilisation of one sub-corpus as the reference to another, functioning as a target corpus, is a common way to analyse which terms are key to either corpus. Given the interest in similarities over simply differences in the present study, however, this approach was rejected following a few attempted analyses, since it became obvious that the approach hid occasions where two disparate discourses used a similar vocabulary. Similarly, using a large general corpus of internet discourse, such as iWeb, was considered but rejected, as it would potentially have hidden recurrences of more ‘normal’ internet discussions and displayed cases of only ‘extreme’ discourse.

Instead, the study relies on the tangentially related Jozef Stefan Institute Newsfeed Web Corpus (JSI-web) (Trampuš & Novak, 2012) – more specifically the 2014-2018 span of this corpus. JSI-web is a monitor corpus of news stories designed to continuously update on a day-by-day basis, gathering solely from internet-based news providers. While arguably counter-intuitive, since JSI does not consist of written peer-to-peer speech or even microblogging

in general, this particular corpus presents period-specific language, functionally acting as more of a period-specific stop-word list than a reference of differences. This solution is somewhat inelegant and future studies should likely rely on a more tailor-made approach.

4.4 Stop Words

Another approach to highlighting keywords in a target corpus is the employment of a list of 'overly common' words which will sometimes clutter the word list and hinder analysis. This approach, known as a 'stop word' list is often frowned upon, as it introduces an element of subjective filtering often performed before any analysis has been attempted, theoretically leading to a loss of potential observations. When I mention above that JSI functions as a form of period-specific stop word corpus, it does so to filter 'current' events of the time, as it has been observed in Zannettou et al. (2018, p 1014) that Gab tends to discuss these events over many other topics. This essentially stops the (key)word list from being cluttered by words pertaining only to period-specific events, allowing more recurring discourse to take its place.

5. Methods

I begin the following section of the paper by briefly presenting #LancsBox, the main corpus tool utilised in the analysis. Following this short introduction, I present, in detail, the methods which are inherent to corpus linguistics and their specific uses in corpus assisted discourse analysis and the statistical measurements of observation chosen from #LancsBox. This latter part can be considered a list of key terms to the approach of the present paper.

5.1 Software

The present study utilises version 4.0 of #LancsBox (Brezina, McEnery and Wattam, 2015), a free piece of software created at Lancaster University. While many corpus linguists still rely on non-corpus-specific software, such as Microsoft Excel, for partial corpus analysis, this approach is mostly limited to small-scale concordance analysis and is cumbersome when handling large corpora. For more advanced types of analysis, a custom piece of software, specifically made for the purposes of corpus linguistics, is needed. Luckily, the availability of software has grown to the point where a website has been made to present all the software according to filtered features and prices (corpus-analysis.com, "Tools for Corpus Analysis"). What makes #LancsBox better than many other free corpus tools, such as AntConc or SketchEngine is that it allows a fair amount of depth when analysing collocations while still allowing for strong visualisation tools. One drawback, however, has been the power of the

tool in relation to the size of the corpus, as the full corpus tends to crash less powerful computers. Luckily, this is mostly solved through working with smaller sub-corpora.

5.2 Concordance

When corpus linguists argue that corpus linguistics are not, in fact, entirely quantitative, the first example will invariably be concordance analysis – a method so core to corpus linguistics that it is, undoubtedly, the most important of all the tools in the toolbox. At its heart, concordance analysis is much the same as any other content-specific analysis; a contextually sound representation of the text, sometimes brought forth through specifically chosen keywords and sometimes randomly sampled. Analysing each of the three sub-corpora requires extensive use of concordance tables, with a slight difference between *Controversial* and *Like*/*Dislike*.

The analyses of both *Like* and *Dislike* repeatedly leverage what is commonly referred to as Keywords in Context, or KWICs, whereby the concordances are filtered based on specific key words or clusters. Since *Controversial* is far smaller, however, keyness and collocation (see below) are largely ineffectual ways of analysis. Instead, observing these concordances is best facilitated by randomising the output based on less directly selective criteria found in the meta-data itself.

Unlike more statistically uniform parts of corpus linguistics, the presentation of concordances is highly variable and dependant on the goal of the analysis. As such, the concordance lines found in the analysis section do not adhere to any specific limitations in terms of word-limits on either side of a keyword but contain as much information as is deemed necessary to use the concordance as an example.

5.3 Keywords

"A word which appears in a text or corpus statistically significantly more frequently than would be expected by chance" (Baker, Andrew, & McEnery, 2006: 97). By contrasting a target corpus against a reference corpus, corpus linguists draw attention to the terms which can be considered 'key' to either. The concept of 'key' refers to terms which are statistically more likely to occur in one corpus over the other and is expressed through what is referred to as a 'keyness' score. Although the terms 'key' and 'keyness' will occur quite frequently in the present paper, I have opted not to include the exact keyness scores in spite of their availability, since the number itself is arguably useless when trying to decipher the meaning of a given observation. That said a quick explanation of its calculation and meaning is necessary.

Using log-likelihood, keyness in #LancsBox is calculated as follows:

$$\text{Keyness} = 2 \times \left(\text{absolute freq} \times \log\frac{\text{absolute freq}}{\text{target corpus}} + \text{ref. corpus} \times \log\frac{\text{ref. freq}}{\text{ref. corpus}}\right)$$

Instead of indicating the results, then, it should be noted here that with a reliability score of 0.01%[1] using log-likelihood, the critical value necessary would be 15.13 to ascertain significance. Unless otherwise noted, all values observed in terms of keyness range considerably beyond this level, with some reaching between 250-1000 keyness score. While the logical assumption would be that a higher keyness score indicate more importance to a term over another, all that actually indicates is that one can say, with 99.9% certainty, that the observed token is more likely to occur in the target corpus in comparison to the referenced corpus.

Like the sub-corpora, following the identification of a word or n-gram as key in the analysis, the discussion of the term as a keyword going forward will be marked/delimited by apostrophes, e.g. 'fake news' or 'alt-right'.

5.4 Collocation

Collocation, or "the characteristic co-occurrence of patterns of words" (McEnery and Wilson 1996, p. 71) considers the relation of a given keyword to any other commonly co-occurring words within a defined sentence span. This is commonly defined as about 5 words to the left and right of the keyword but is highly variable depending primarily on the goal of the analysis. According to Baker (2006, p. 96), frequent collocation of two words lends evidence that the discourses surrounding them are 'particularly powerful', suggesting that further qualitative analysis is required. In the present study, collocation is particularly useful in ascertaining the words most commonly associated with news institutions and may for example highlight co-occurring hashtags in addition to common lexical occurrences. Unless otherwise stated in any part of the analysis, the default span has been kept to 5 words left and right of the search term as the limitations of any collocation analysis.

Unlike keywords, the strength of collocations is measured in terms of 'mutual information' (henceforth, MI), or on some occasions the relative expected occurrence measured in confidence through a t-score. However, since even the most focused corpora here are relatively large and I choose to deal only with the highest ranked collocates, MI will suffice. MI is a term which expresses the degree to which a word tends to co-occur with another word, within the given span. In other words, in a given corpus MI is calculated on the basis of the number of times a pair has been observed together, versus the number of times the pair has been observed apart. As an example, the word which most strongly collocates with the node

[1]

word ‘news’ in the *Like* corpus is ‘fake’. ‘News’ occurs 557 times and ‘fake’ occurs 218 times as a collocate to ‘news’ out of the 295 times it occurs altogether in a corpus measuring 481,494 words. The mutual information score here, which is calculated logarithmically in #LancsBox, is thus handled as follows: $\mathrm{MI} = \log_2 \frac{co-occurrence}{node\ word}$, here resulting in an MI score of ~ 9.31. Like the calculation of keywords, MI itself only measures the certainty with which one can say these two words are likely to occur, as opposed to them occurring individually, outside of the pairing. In the present study, the threshold for accepted co-occurrence in #LancsBox has been set to ‘5’.

Beyond calculation, #LancsBox allows for positional collocation mapping, which indicates where, in context to the head word the collocate is placed. To give an example, figure 2 shows the collocations of the keyword ‘example’ in the *Like* corpus. Beyond simply highlighting the collocates themselves – words such as ‘is’, ‘of’, ‘an’ and ‘the’ – the saturation of the node displays the strength of the collocation, abstracting the MI score, while the position of the node relative to the keyword presents the recurring position of the collocation in context. Looking at ‘of’, for example, suggests its use as ‘example of’, as it tends to appear to the right of the keyword, while ‘an’ and ‘perfect’ appear exclusively on the left. From the position of ‘perfect’, right next to the keyword, it is also apparent that it appears close to the keyword, presumably as ‘perfect example’, as opposed to ‘in’ which tends to occur further away, although still within the span of 5 words left/right of the keyword.

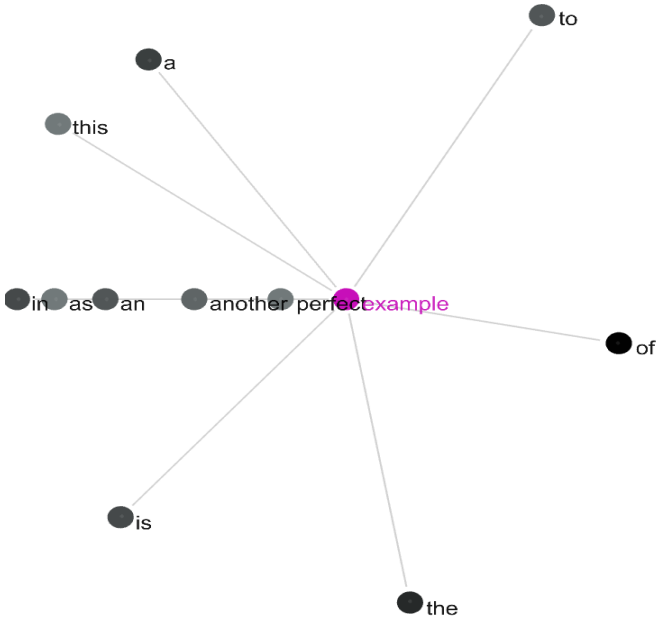

Figure 2. Example of collocate network of ‘example’ in #LancsBox

5.5 Clusters

Clusters are defined by Baker, Hardie and McEnery (2006, p. 34) as "any group of words in sequence", forming a helpful picture of the multi-word terms inherent to a particular discourse, or sometimes the types of sentences a single-word term is commonly used in. Unlike the open-ended nature of collocation, where any co-occurrence within the applied sentence span is registered and tallied, clusters must recur in exactly the same form multiple times to be considered. While this limits the results, it emphasises instantiations of particular clustered expressions, such as "fake news network" or "hate crime". Together with collocation, then, cluster analysis provides a view of the types of prosody inherent to a given word or phrase. In the analysis itself, clusters are commonly referred to as n-grams, where 'n' signifies the limit to the amount of words contained within the cluster. The analysis of the present paper deals almost exclusively with two-word n-grams, referred to simply as bi-grams.

6. Analysis

In an effort to bring overview, the following section details the logic behind the development throughout the analysis.

The response to the criticism levelled at Gab, from spokesperson and site-creator Andrew Torba, has been to argue that not only is Gab not a safe-space for white supremacy or white nationalists in particular, Gab includes everyone and spans a wide spectrum of user-groups and viewpoints. To ascertain the validity of the claim that Gab represents a varied and wide user-group, I analyse the *like*, *dislike* and *controversial* sub-corpora, highlighting the top keywords and following up with detailed collocation and concordance analysis of particularly popular sentiments.

Where possible, each piece of the analysis will follow a similar formula, moving from qualitative observations to quantitative and back to qualitative once more: I begin by isolating keyword clusters according to their relative frequency in the corpus and present these as lists. I then evaluate the importance of the sentiments and, where applicable, combine semantically consistent occurrences. Finally, I then take these individual sentiments and make more in-depth inquiries as to their usage through qualitative analysis of the concordance lines, highlighting their use in context.

6.1 [De]legitimation

One of the most prevailing critiques levelled at Gab has been that it functions as an echo chamber, with little space for those who do not consider themselves part of the alt-right, alt-

lite or simply those who openly reject intense nationalism and neo-conservatism. The site creators, headed by Andrew Torba, have been vocal in denying such criticism. According to them, Gab is for everyone, and the users span a multitude of geographic, ideological/political and racial spectrums.

Ascertaining the sentiments of an entire community is arguably more than what can be handled by qualitatively analysing single pieces of Gab discourse. To test this assumption, then, the following analysis attempts to discern differences in opinion amongst the most upvoted, downvoted and controversial statements, to offer some insight into the discourses and topics which are legitimised and de-legitimised by the community. The method with which each corpus was created is discussed earlier in section 4.2.

6.1.1 *Like*

Like is the largest of the three sub-corpora, consisting of 32,880 posts spanning just over 276,000 words. To ascertain discourse patterns, I begin by looking at simple keyword clusters, using the JSI Web Corpus (2014-2018) as my reference corpus. Using #Lancsbox, the n-gram has been set to allow a span of up to 4 words. Unsurprisingly, however, all the most recurring ones appear as bigrams.

N-grams Keywords in *Like* ranked by Freq per 100,000 words (ref. JSI Web Corpus 2014-2018)

Rank	**Freq**	**Keyword | Single Word**	**Rank**	**Freq**	**Keyword | N-grams**
1	91.3	free speech	11	7.6	terrorist organization
2	39.5	fake news	12	7.6	terrorist attack
3	22.5	hate speech	13	7.6	white supremacist
4	12.7	western civilization	14	7.2	american flag
5	12.7	civil war	15	6.9	illegal immigration
6	12.3	deep state	16	6.5	voter fraud
7	11.6	3rd world	17	6.5	safe space
8	10.5	travel ban	18	6.2	terror attack
9	9.8	political correctness	19	**5,8**	red pill
10	8.7	hate crime	20	5.5	white genocide

Table 2. N-gram Keywords in *Like* ranked by freq.

When analysing keywords, it is customary to isolate particularly interesting occurrences from the list, in order to further investigate their meaning in context. I would argue, however, that no such fine-graining is necessary with the results found in Table 2. while certain occurrences, such as ‘free speech’ and ‘political correctness’ are somewhat expected, in light of the site’s inherent focus on the domain of free speech, the same can hardly be said for terms such as ‘deep state’, ‘red pill’, ‘white genocide’, ‘race war’ or really any other term found therein.

Additionally, terms such as ‘voter fraud’ and ‘illegal immigration’ mirror many of the keywords often expressed by conservative news institutions such as Fox News and Breitbart, according to earlier investigations by Christiansen (Unpublished).

To create some semblance of structure, then, the analysis going forward will deal with three separate aspects highlighted as agreed-upon in the community based on their recurrence as keywords in the ‘like’ sub-corpus. The first incorporates the 1st, 3rd and 10th highest ranking n-grams, ‘free speech’, ‘hate speech’ and ‘hate crime’, with the aim of ascertaining the community’s definition and understanding of these. The second deals with the 2nd highest ranked n-gram, ‘fake news’ and aims to investigate what the community defines as ‘fake news’ versus trusted news. Lastly, the third aspect is the most abstract, including the 4th, 5th, 6th, 13th, 19th and 20th highest ranked n-grams, these being ‘western civilisation’, ‘civil war’, ‘deep state’, ‘white supremacist’, ‘red pill’ and ‘white genocide’. This last section includes all terms directly and almost uniquely related to the alt-lite and alt-right and may help in determining the attitude of the community towards these groups.

1. ‘Free speech’, ‘hate speech’ and ‘hate crime’

Beginning with ‘free speech’, shown in table 3, the most frequently recurring collocates seem to fall within two categories; discussion of the website and its mission, shown by ‘Gab’, ‘Online’, ‘Platform’ and ‘Internet’, and a prosodic sense of danger towards ‘free speech’ as a concept, through censorship and other opaque enemies, as well as calls for support for, or statements of defence of, free speech.

Collocates of 'free speech' ranked by MI

MI	Freq	Collocate	Position
8.1	8	@gab	L
8.0	5	protection	L
7.6	15	platform	R
7.6	5	individual	R
7.5	5	liberty	R
7.2	15	defend*	L
7.0	19	support*	L
7.0	5	alt	L
6.9	5	rally	R
6.6	7	rights	R
5.7	7	banned	L
5.1	6	attack	M
4.5	7	against	L
4.0	72	is	R
4.0	6	hate	L

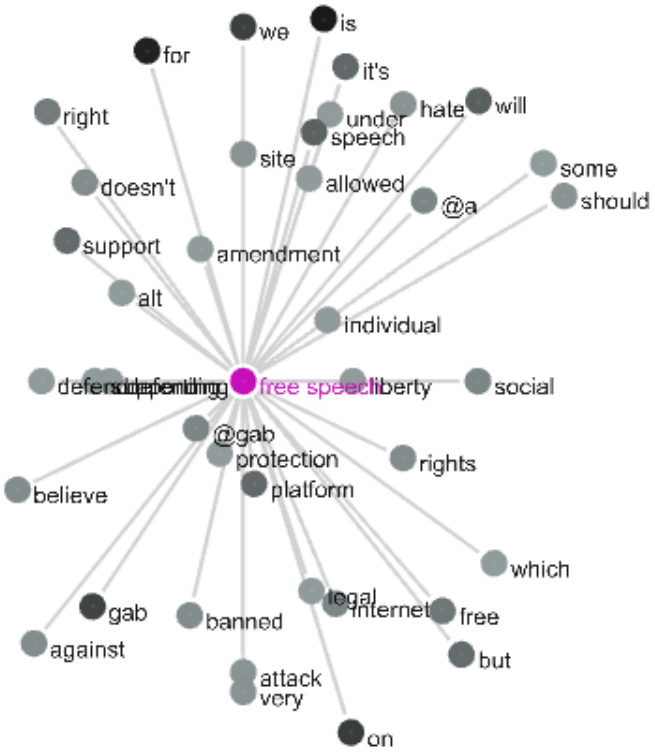

Table 3. Collocates of 'free speech' in *Like* ranked by MI.

Apart from an apparent collocative connection to 'free speech', however, little can be surmised from the results in 'hate speech', as seen in Table 4.

Collocates of 'hate speech' ranked by MI

MI	Freq	Collocate	Position
7.5	5	isn't	R
6.6	8	speech	L
6.0	5	down	L
6.0	5	free	L
4.6	24	is	R
4.5	15	for	L
4.5	7	not	L
4.2	6	with	L
4.1	9	this	L
3.8	7	on	R
3.8	23	to	L
3.7	6	that	L
3.6	13	and	L
3.2	11	a	R
3.1	20	the	L

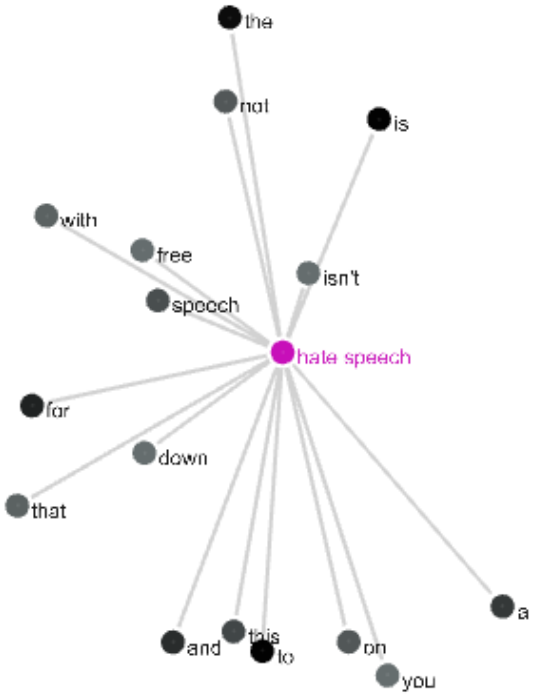

Table 4. Collocates of 'hate speech' in *Like* ranked by MI

The most obvious way to analyse the community definition of the terms, however, is through targeted collocation analysis, specifically pairing each of the terms with the appropriate form of the verb 'to be' and then analysing concordances. Essentially, this isolates instances where people have been supported in their views of what 'free speech' and 'hate speech' both is and isn't. In this case, the collocational values back this up as well, as both 'hate speech' and 'free speech' collocate strongly (in terms of MI) and frequently with 'is'. Keep in mind that although 'is' has relatively low MI scores in both, this is due to its extremely high occurrence in the general English language.

Concordances of 'free speech' collocating with 'is' in *Like*

		Keyword	
1	Everyone won't agree with you. **Hate speech is**	free speech	5. Facts over feelings
2	Google is **cracking down** on	free speech	but patriots are standing up
3	Anyone telling you they'd die for your right of	free speech	is full of shit. People in modern society wouldn't
4	Our democracy and our	free speech	**is just a staged show**
5	I'm enjoying this**. It's the wild west of**	free speech	. This is the place to be.
6	They are **walking the line of**	free speech	, but it is indeed free speech. We are monitoring
7	Gab is **the last home of**	free speech	on the internet. I have effectively been banned by
8	I'm for FREE SPEECH. **The above is NOT**	free speech	. Get the hell out.
9	In America we always talk about how	free speech	**is dying (it is)** but we often don't realize how
10	You all need to get on the DS Forum. It is **the last place for**	free speech	on earth.
11	Weird to think we live in an age where **defending**	free speech	**is considered an abhorrent act** and killing babies is considered freedom.
12		free speech	is bad, illegal immigration is good. Welcome to 2017.
13	This is a First Amendment site.	free speech	**goes**. Twitter has mountains upon mountains of racism towards white people.
14	20k avg Gabs shared per day.	free speech	**is cool again**.
15	**I just love** liberty, freedom, and most of all	free speech	. The rest is just noise

Table 5. Concordances of 'free speech' collocating with 'is' in *Like*

Looking at the entire collection of collocates, the most recurrent sentiment centred around the concept of free speech in the 'like' corpus is that free speech is *under attack*. This is often reflected in largely opaque mentions of having to 'defend free speech', from loosely defined institutions ranging from Twitter to the EU, CNN, Socialists, The Elite, Big Social, Liberals and Jews. Sometimes Gab itself is even criticised for not defending free speech well enough. These mentions, exemplified in Table 5, lines 2, 4, 7, 9, 10 and 11 amount to more than 40 percent of the sentiments expressed towards 'free speech', suggesting that it is both a recurrent and well-liked sentiment. Analysing the sentiments individually, 'free speech' as a concept is presented as a moral or ethical code to be followed, and companies or individuals who obstruct or limit it in any form are viewed as immoral and socially sanctionable. This

observation is further supported by occurrences of people discussing the malleable nature and limits of free speech, as is also seen exemplified in lines 6 and 8.

In an example such as L2, "Google is cracking down on free speech, but patriots are standing up" this supposed attack is further used to legitimise the opposition to google as 'patriots' of free speech. The discourse also often exhibits unclear presuppositions, such as in L14, where 'cool again' insinuates that free speech had at some point fallen out of favour, or L9, where the speaker simply claims, on unclear grounds, that free speech is under attack. Numerous examples in the corpus also see people present free speech as having some form of successful final defence.

Concordances of 'hate speech' collocating with 'is' in *Like*

		Keyword	
1	Google: Gab needs to **police**	"hate speech"	. Also Google: Gab can purchase ads from Google and target "hate speech".
2	the Met police were too busy trying to **hunt down** all of the bigots posting	hate speech	online
3	Wow, but	"hate speech"	is the **"real problem"** in Australia I guess?
4	Use this topic to share	"hate speech"	from other social networks
5	#Netherlands politician convicted of	hate speech	under the **guise** of "inciting discrimination".
6	A massive win for free speech! Bookmark this for anytime someone brings up	"hate speech"	.
7	#Snowflake EU is threatening social media networks, if they don't take steps to **censor**	hate speech	within 24 hours. Yet they still wonder why the citicizens of #UK voted to leave
8	When they put 'free speech' in quote marks, or **label** non-leftist speech as	hate speech	, you know they are trying to infringe upon the first amendment
9	Had Trump lost, the Left would be bearing down on America with **totalitarian**	'hate speech'	**legislation**, ready to go.
10	(This is unbelievable): apparently some retarded liberal **reported** me for	'hate speech'	because I called him retarded. This after calling me a 'Trump Nazi' BTW.
11	Pro tip: there is no line. Free speech includes	"hate speech"	.
12	you understand that "hate speech" is not a legal concept, and that what people call	"hate speech"	**is protected** under #1A, right? Cause this is basic.
13	That and continue to suffocate our freedom of speech with **bull shit**	"hate speech"	laws/**shadow bans/censorship**.
14	Know your rights! Hate Speech is Bad! Down with	hate speech	!
15	Whites aren't evil. Frogs aren't racist. Islam isn't peaceful. **Truth isn't**	"hate speech"	. Feminism isn't "equality"

Table 6. Concordances of 'hate speech' collocating with 'is' in *Like*

Not surprisingly, the sentiments levelled at 'hate speech' bears almost no resemblance to those expressed in relation to 'free speech'. Particularly interesting here is the highly frequent use of quotation marks or inverted commas surrounding the term. From context, it is clear that this is done with two interdependent purposes: on the one hand, it likely serves as a counter-discursive measure, allowing users to paraphrase what they perceive as the popular

definition of hate speech. On the other, it is used to mock those who would use the term seriously, often with the reasoning that ‘hate speech’ is an attempt at silencing those with controversial opinions or people trying to speak “truth” to power. Ironically, second usage is perhaps best exemplified in L8, where a user is lamenting the use of quotation marks in describing ‘free speech’, but it can be seen in almost every single line in table 6.

Out of all the sampled concordances found in table 6, however, only a single person is lamenting the recurrence of hate speech itself, although the post itself betrays very little as to the context of the lamentation. Conversely, the use of quotation marks to imply sarcastic usage means that while the majority of collocations with ‘hate speech’ are negative, such as the verbs ‘police’ and ‘hunt down’ or ‘report’ this does not express a negative attitude towards the type of hate speech defined by Holbrook’s extremist media index, but rather shows negative attitude towards the definition itself.

Concordances of ‘hate speech’ collocating with ‘is’ in *Like*

		Keyword	
1	**WTF IS A**	"HATE CRIME?"	If you kidnap, attack, torture, or assault someone, it's a felony. End of story.
2	WOMAN **ADMITS TO FAKING** #TRUMP	'HATE CRIME'	'I was suffering from depression at the time (…)
3	of a 'special needs' white man for 24-48 hours.... **not sure if this is a**	'Hate Crime'	it may just be' Young kids doing something stupid'- You can't make the shit up
4	Gun Crime UP 27% (despite draconian gun control!) > Total Crime UP 19% BUT **arrests for**	"Hate Crime"	**(by the Thought Police) UP 53%.** That's for "offensive comments online".
5	SHARE WIDELY!!! We demand	hate crime	charges IMMEDIATELY.
6	study has found that left-wingers and Muslims are responsible for the overwhelming majority of **anti-Semitic**	hate crime	in several different European countries
7	Muslim extremist who **lied about**	hate crime	goes into hiding to avoid being arrested for hoax charges.
8	know about the rest of you, but I cannot even bring myself to watch the	hate crime	video. Just the pictures make my heart race and bring me anxiety.
9	**If giving a Muslim bacon is a**	hate crime	due to it being offensive to their religious beliefs, then forcing a Christian to accept homosexuality and transgenderism is also a hate crime
10	another day **another**	hate crime	**hoax.**
11	Black liberal reporter gets 5 years in prison for **staging**	hate crime	**hoaxes**. Finally, some of these leftist dirtbags are being punished!
12	French Mayor To Be **Charged With**	Hate Crime	For Decrying Replacement of White People
13	We’ve received many enquiries as to what constitutes a	hate crime	. There’s really no set standard, it’s just whatever we feel like it is.
14	The Black Lives Matter Kidnapping is **a verified racist**	hate crime	, so why is Tucker Carlson the only MSM reporter currently covering it?

15	NO! WTF is this shit?!? **This is a** fcn	hate crime	! It's a bunch of blacks recording the torture of little White girls.

Table 7. Concordances of 'hate speech' collocating with 'is' in *Like*

Despite 'hate crime' not taking on the same quotation-mark-indicated sarcastic presentation as 'hate speech', the terms still mirror each other in a number of ways. First, the legitimacy of 'hate crime' as a concept is questioned in much the same way, as is perhaps best exemplified by someone mockingly joking that there is no standard for 'hate crime', "it's just whatever we feel like it is" in Table 7, line 13. This sentiment is repeated in lines 1 and 4, towards an argument that the concept of 'hate crime' is used to silence, in the same way as 'hate speech', exemplified in line 12. Finally, the ultimate form of this sentiment is expressed through the sanctioning of 'hate crimes' as either "fake" or "hoaxes", as seen in lines 2, 4, 7, 10, 11.

The only occasions where hate crimes are portrayed as a legitimate concern, according to the most liked sentiments, are when they are committed by minorities and the political left and when the targets are specifically white. This is expressed through outrage over particular scenarios involving white victims, such as the supposed "Black Lives Matter Kidnapping" (see e.g. Menegus 2017) referenced in lines 3, 5, 8, 14 and 15, and more generally by referencing situations where minorities have been the arbiters of hate crimes, as seen in lines 6, 7 and 11. Lastly, frustration with the mainstream media's representation of events, which can be seen throughout the corpus itself, is on display in line 14 and 15, with 14 especially expressing wonder at Fox News' Tucker Carlson being the only anchor reporting on this view of the incident.

2. 'Fake news'

The frustration with mainstream news and portrayal of the 'msm' as dishonest or unwilling is perhaps best exemplified by yet another key term on Gab. One of the most popular topics amongst the alt-lite and alt-right is the denigration of 'fake news', and its appearance as one of the most recurrent keywords in the 'like' corpus is hardly surprising to those who would call Gab a fringe-right community. To ascertain the community's stance on what counts as 'fake news', I begin by analysing the top collocates.

Collocates of 'fake news' ranked by MI

MI	Freq	Collocate | Single Word	Position
7.9	9	narrative	R
7.3	22	CNN	L
7.1	18	propoganda	R
7.1	17	fake	R
6.8	7	press	R
6.5	7	msm	L
6.2	6	#pizzagate	L
6.0	15	news	L
5.9	8	very	L
5.8	5	russia	R
4.5	22	they	L
4.4	15	Trump	L
4.3	8	it's	L
4.1	48	is	L
4.0	6	more	L

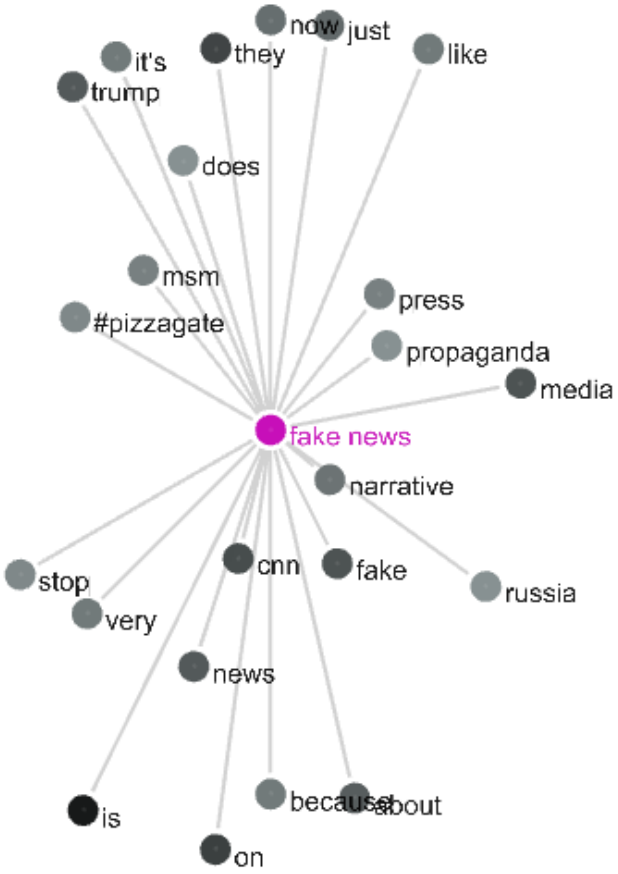

Table 8. Collocates of 'fake news' ranked by MI

From these collocates alone, it is possible to derive a fair amount of meaning regarding the discourse prosody surrounding the concept of 'fake news'. Words such as 'narrative', and 'propaganda' in connection to institutional identifiers such as 'CNN', 'msm' (mainstream media) and 'press' highlights the focus on storytelling and the attitude towards particular institutions as producing 'fake' stories for the purposes of propaganda. Once again, fake news is used to identify certain targets, made clear through the term's relatively strong collocation with forms of 'to be' in 'is' and 'it's'.

Concordances of 'fake news' collocating with 'is' in *Like*

		Keyword	
1	Reality: All are welcome on Gab	Fake News	: Gab is for Nazis.
2	**Wikipedia is** a	fake news	source. Infogalactic – which will displace Wiki – is accurate
3	NBC's Fake News King Brian Williams Launches Crusade Against	Fake News	. What is that saying about glass houses and stones?
4	And **they tell me that #Pizzagate is**	fake news	. If that is true, why you making so much noise, Clinton? Something you want to tell us?
5	About a week ago, **CBS declares on 60min that Mike Cernovich is**	FAKE NEWS	CBS knows fake news- because **they ARE fake news**
6	The MSM says **#Pizzagate is**	"Fake News"	. They said the same thing about Breitbart when they were the only news outlet investigating Anthony Weiner

7	**Everything they write about Russians and Trump is**	Fake News	. Wise up to Deep State nonsense.
8	As if we didn't know all along, because **CNN is**	Fake News	.
9	**CNN is** the worldwide headquarters for	Fake News	.
10	Will Snopes debunk the story of their co-founder embezzling money to pay for hookers? Maybe they can **just label it**	"fake news"	and wait for it to go away. #speakfreely #MAGA
11	Each time the media refers to the US as a "democracy"**, it's**	fake news	. Ours is a constitutional republic and it protects us from the tyranny of the majority.
12	**the so-called "mosque attack" is** looking increasingly like	Fake News	. /pol/ is on it. Check out the pictures
13	NEW MEME:	Fake news	**is a "moral panic."** In other words, it's a trivial problem being blown way out of proportion. Spread this meme on Twitter.
14	**Not**	Fake News	NOBODY is Allowed To Talk About The Sick Things these Muslim Refugees did […] Judge Gagged EVERY-ONE including her Own Parents ~!!
15	**Facebook is so focused on** their	"fake news"	problem, meanwhile folks are live streaming murders.

Table 9. Concordances of 'fake news' collocating with 'is' in *Like*

While the popular discourse on 'fake news' is quite varied, it tends to be similar in nature to the duality found in 'hate speech' and 'hate crime'. The term is heavily used to embellish or discredit institutions which have been deemed illegitimate by the community, such as Wikipedia, CBS, CNN, NBC and the broader "MSM" (mainstream media) as is seen in lines 1, 3, 5, 7, 8 and 9. At the same time, the use of term to accuse the media of creating a "moral panic" is exemplified in lines 10, 13 and 15. This approach shows similar tactics to those found in 'hate speech', where users are attempting to delegitimise the use of the term 'fake news' by what they deem illegitimate institutions or individuals, essentially arguing that people are abusing the term or are, themselves, "fake news". Things which are deemed to be *not* fake news include the popular far-right conspiracy theory "Pizzagate", shown in lines 4 and 6, as well as a story supposedly about Muslim refugees in line 14 and Mike Cernovich, a prolific alt-right spokesperson in line 5. The recurrence of these items, in particular 'pizzagate', which also shows up as one of the top collocates in Table 8, reflects the connection of the term 'fake news' to the broader sphere of alt-right and far-right discourse. Its recurrence in *Like*, of course, reflects the popularity and legitimation lend to the theory on Gab in general.

3. 'Western civilization' against 'white genocide', and 'civil war' – why we must all be 'red pilled' from the 'deep state'

Concordances of 'western civilization' collocating with 'is' in *like*

		Keyword	
1	You're not going to save	western civilization	sitting on the couch. Hit the weights. Get in the gym. Antifa doesn't even lift.

2	We need more militancy and less fagotry on the Right.	Western civilization	is literally on a path to extinction
3	It is a religion of hatred A religion of bloodlust **A religion that is incompatible with**	Western Civilization	Pray for the victims Pray for Europe to wake up
4	The United States is a White, Christian nation. We are an extension of **Christendom and**	Western Civilization	. We are not a "Judeo" nation. **Jewish values are incompatible** with American values.
5	Dear **Black People,** You **are asking to dismantle**	Western civilization	. I think we'll pass.
6	**You guys built the backbone of**	Western civilization	, advanced progress, modernity, The Enlightenment and the autonomy of man.
7	**Toxic masculinity is what built**	Western civilization	. It's what builds most things**. Feminism does nothing but destroy**
8	Get over yourself.	Western civilization	is under attack. We must fight.
9	Paul Joseph Watson: It's about reinvigorating the American Dream. It's about ultimately saving	Western Civilization	. CNN: All of this is very disturbing for many people in the rest of the political spectrum.
10	**The greatest threat to**	Western Civilization	**is now the cuckservatives**. Time to turn up the heat on these timid cowards.

Table 10. Concordances of 'western civilization' collocating with 'is' in *Like*

Like 'free speech', 'western civilization' is a continuously legitimated institution, one which has been built by 'man', as shown in lines 6 and 7 and which reflects 'white' values, 'christendom' and good Christian values in general. Also like 'free speech', 'western civilization' is discursively positioned as being constantly under attack from everything from "feminism" (L7) to "antifa" (L1), "black people", (L5) the Muslim faith (L3) and "judeo" values (L4). Interestingly, 'western civilization' has an enemy not discursively present in 'free speech': the so-called 'cuckservatives', seen in lines 1 and 10. The term 'cuckservative' is a pejorative used to describe a conservative voter who does not align properly with alt-right/alt-lite values, refusing the militancy referenced in line 2, and consequently appearing 'timid' as noted in line 10. It is also worth noting that, in order to 'save' western civilization, users are being told to prepare for a fight in lines 1, 2, 8 in a very material sense, with reference to training in line 1 and militancy in line 2.

Concordances of 'white genocide' and 'civil war' in *like*

		Keyword	
1	Normie whites: "There's no such thing as	White Genocide	you racist Nazis"
2	Sex war? Intelligence war? Digging up graves?	White genocide?	Alt-Left: the face of hate.
3	**Once they admit** that	white genocide	is happening, they say it doesn't matter because it's a good thing. Typical.
4	Any wonder why Hillary had him executed? Ya, but this whole	white genocide	**thing is a crazed conspiracy theory right?** @Shatteredkarma

5	"There is no place in modern Europe for ethnically pure states"	White Genocide	? **There is no such thing, right?** Did anyone ask your opinion about this?
6	In a hypothetical	civil war	, just imagine for a moment if you will **the emasculated men of the left leading the resistance**.
7	The left would rather risk nuclear war with Russia &	civil war	in America than accept Trump's victory. Incredible.
8	Trump is inciting civil war! Says CNN, who have endorsed alt-left groups that specifically call for	civil war	.
9	**Red pill: we're in a cold**	civil war	and the future of the West depends on who wins the culture, tech and media battles. Blue pill: everything is fine.
10	Dear Senators, I hope you understand how close we are to	civil war	if you try to overthrow our President. For your sake, get the message.

Table 11. Concordances of 'white genocide' and 'civil war' in *Like*

The vague explanation of this war on 'western civilization' seems to be strongly associated with the notion of an impending 'civil war', as seen in lines 6-10 and resulting 'white genocide' as seen in lines 1-5. Contextually, 'civil war' and 'white genocide' often present an interesting paradoxical duality: On the one hand, users boast how easy it is to beat their enemies, often due to their physical weakness (see e.g. Table 10, line 1 and Table 11, line 6), but on the other, there is strong support for the sentiment of victimisation. This victimisation comes through 'white genocide' a conspiracy which suggests that white people are rapidly losing out to other ethnicities and becoming a minority in western countries, often due to leftist politics. This victimisation is, itself, somewhat paradoxical, seemingly dependent on a war which is simultaneously about to get started (e.g. Table 11, lines 6, 7 10) and yet half-way lost by white people. In this regard, it seems that the users consider themselves part of a sort of elite which has seen the light and is therefore fighting, in stark contrast to those who are victims but are not fighting, i.e. the 'cuckservatives' and the 'sheeple', as shown in e.g. Table 10, line 10.

Concordances of 'red pill' in *like*

		Keyword	
1	Reminder to homeschool your kids, raise them in God's country not the degenerate city, and	red pill	them on our corrupt, wicked culture as soon as they start communicating
2	Delicious!!	Red pill	pusher @PrisonPlanet has made BBC News
3	Someone please	red pill	me on this.
4	You are more than welcome here on Gab. We want you. We appreciate you. Let's	red pill	the masses, together. Positive vibes.
5	This is not surprising to many of us, but needs to be spread. Big	red pill	here for CNN sheeple.

Table 12. Concordances of 'red pill' in *Like*

The process of transforming from a ‘cuckservative’ into a more legitimate member of the community, at least according to Gab, requires one to be ‘red pilled’ through explanations of government and, as shown later, the broader conspiracy of the ‘deep state’. The concept of the ‘red pill’ comes from a famous scene in the 2000s movie *The Matrix*, where the main character, Neo, is asked to choose between a red pill signifying awakening and acceptance that things are bad, in order to make a change and a blue pill which allows him to ‘go back to sleep’ and forget everything. As such, ’red pill’ has inherently negative connotations, in the sense that it always presents the ugly truths of the world but is necessary since it allows one to see the world as it truly is. Legitimate institutions connected to the concept include Prison Planet – a website with a heavy focus on anti-democrat content, rated as “conspiracy/pseudo-science” by MediaBiasFacts.

Concordances of ‘deep state’ in *Like*

		Keyword	
1	The OKC bombing ended the militia move-ment. I guarantee the	(((deep state)))	is working on 2.0 to destroy the AltRight
2	Top troll Also sends a hidden message to the media, the	deep state	, and the alt left: you will be brought to heel.
3	After months of seeing how	Deep State	had its own agenda, we are asked to take its word on Syrian “gas attack”.
4	The Flynn resignation is part of a concerted	deep state	effort to undermine, infiltrate & sabotage Trump.
5	I’m guessing Deep State works best when most ppl are unaware that	Deep State	exists. But now Deep State has pulled its cock out and is waving it around.

Table 13. Concordances of 'deep state' in *Like*

Alongside all of the agitators mentioned in relation to the war on ‘western civilization’, one group of people is at once opaque and more ubiquitous than any other; the illusive ‘deep state’. While this term in itself betrays an allegiance to alt-right conspiracy theories, mirroring the language of Infowars and Breitbart, it is only further exacerbated by the use of the triple parenthesis around ‘deep state’ in line 1. The triple parenthesis, known as an “echo”, is a tool used by white nationalists to imply hidden Jewish involvement. In 2017, the Anti-Defamation League officially declared this demarcation a general hate symbol (ADL, “Echo”).
In summary, none of the words which are commonly related to the alt-right or white nationalist beliefs are used in discussion alone or with sarcastic intent. While there are no outright statements encouraging violent action, the terms tend to relate to a discourse of antagonism and othering in a more extreme fashion than is seen in both ‘fake news’ and ‘free speech’ / ‘hate speech’. Through the recurring construction of the western world as being ‘under

attack' and even in the process of genocide, further it is sometimes directly indicated that this is due to Jewish involvement.

6.1.2 *Dislike*

Dislike is the second-largest of the three sub-corpora, consisting of 23,000 posts spanning 216,800 words. As with 'like', I begin by looking at simple keyword lists, once again using JSI Web Corpus (2014-2018) as my reference corpus. Early efforts at analysis using single-word keywords made it clear that some of the most consistently downvoted posts and users were spam posts, attempting, through trolling, to provoke other users on the site. These posts, which often consist of simply the word 'haha' or 'lolol' repeated to various lengths, are of less interest to us and have therefore been manually removed. What is interesting, however, are the posts which, while trolling in nature, take a clear stance towards particular people or institutions. While it can arguably be difficult to differentiate between occasions where a post has been downvoted due to its content rather than its trolling attitude, these posts have been kept, as downvoting of clear stance-taking indicate political attitude and are thus of interest.

N-gram Keywords in *Dislike* ranked by Freq per 100,000 words (ref. JSI Web Corpus 2014-2018)

Rank	Freq	Keyword | Single Word	Rank	Freq	Keyword | N-grams
1	44.3	real life	11	8.3	|X| lol
2	25.8	free speech	12	7.8	christian account
3	12.1	|x| She	13	7.4	personal property
4	10.6	sick fuck	14	6.9	old slag
5	10.1	real name	15	6.9	flat earth
6	9.7	child molestation	16	6.5	right wing
7	9.7	sex offender	17	6.5	fake news
8	9.2	nazi trash	18	6.5	safe space
9	9.2	hate speech	19	6.5	evasive chickenshit
10	8.3	bedpan licker	20	3.7	bot control

Table 14. N-gram Keywords in Dislike ranked by freq.

Looking at the top ranking n-grams found in the 'dislike' corpus, it is particularly interesting to see certain key terms, such as 'free speech', 'hate speech', 'safe space' and 'fake news' occur in both the like and dislike lists. This recurrence warrants further investigation into the differences in attitude presented by posts which are deemed legitimate versus non-legitimate by the community. It is also interesting to see the female pronoun 'she' at the start of the sentence, indicated by |x|, warranting an examination of female representation within delegitimised posts. Finally, another key characteristic of the downvoted posts has a strange wrinkle to it in name-calling, with terms such as 'sick fuck', 'nazi trash', 'bedpan licker', 'old slag'

and ‘evasive chickenshit’ showing as particularly key. While name-calling itself may indicate attitude, the essence of which is here obviously aggressive and negative, it once again serves as a statistical oddity. Initial analysis showed that while the posts circumvented the de-duplication process by not being exact replicas of each other, and while some of them were made by more than one account name, it seems extremely likely that these come from a single user attempting to spam the same comment over and over again. This observation includes all occurrences except for ‘nazi trash’, and since that particular occurrence is fairly obvious in the people it targets, further analysis is arguably unnecessary.

To split the analysis into groups in the same was as *Like*, then, the continued analysis of *Dislike* deals with two separate aspects and four key occurrences. The first deals with recurrence, including three separate observations under a common umbrella; ‘free speech’, ‘hate speech’ and ‘fake news’ all of which appear as particular key to both *Like* and *Dislike.* The the conflicted use of these terms further highlights the differences between discourses which are deemed legitimate and illegitimate by the community. The second examines the representation of ‘she’ in *Dislike* and briefly mirrors this to its representation in *Like*

1. ‘Free speech’, ‘hate speech’ and ‘fake news’

Given the relatively low frequency of the term ‘hate speech’ in the dislike sub-corpus, collocation analysis is arguably unnecessary, but a few interesting results do pop up from the relatively more frequent ‘free speech’.

Collocates of 'free speech' ranked by MI			
MI	**Freq**	**Collocate**	**Position**
11.1	14	encourages	L
10.9	6	wrecked	L
10.8	14	behave	L
9.1	14	anonymous	L
9.1	6	SJW	M
9.0	18	platform	R
7.6	13	speech	R
7.4	13	hate	R
7.1	14	accounts	L
6.9	6	Antifa	R
6.1	6	#maga	R
5.9	18	people	R
5.6	25	not	M
4.6	6	gab	L
4.1	23	**is**	L

Table 15. Collocates of 'free speech' in *Dislike* ranked by MI

Initially, many of the key collocates here appear to mirror the ones found previously, with terms such as 'gab', 'SJW' and '#maga' reflecting an extremely similar discursive prosody to the one seen in the *Like* corpus. Even words such as 'antifa' has previously appeared as negatively evaluated institutions in relation to 'free speech'. A quick look at the concordances, however, clearly shows that the context lend to these occurrences in *Dislike* is quite different. Once again, I begin by observing the collocations of 'free speech' with 'is', based at least partially on its occurrence in the table above.

Concordances of 'free speech' collocating with 'is' in D*islike*			
		Keyword	
1	you are a fool or a shill if you think @a is for	free speech	.
2	**hate speech is not**	free speech	#speakfreely but kindly
3	#AltRight is in ruins. Milo's downfall may just be the end. **Hate speech is not**	free speech	. This is long overdue.
4		Free speech	**is an enlightenment principle**. Never mentioned in the Bible.
5	Gab is awesome because it's	free speech	for everyone, alt-right to Antifa.
6		Free speech	for all including Antifa.
7	Looks like Gab opened the flood gates for more deplorable racist, anti-Semitic users. Shocking. Why don't you go back to Twitter where there is	free speech	with not hate speech?
8	Where is my	free speech	, hypocrite?

9	I meant **hate speech is not**	free speech	you dipshit
10	Hillary didn't kill Seth Rich, and no one cares. You guys are pathetic, but that's my opinion so it can't be deleted, right?	Free speech	**is the bomb**.

Table 16. Concordances of 'free speech' collocating with 'is' in *Dislike*

As a sort of directly mirrored reflection of the values expressed in the *Like* corpus, the discursive representation of 'free speech' in *Dislike* shows almost directly opposing opinions, highlighting the degree to which certain values and discourses are not only legitimated through support, but also through delegitimating opposing opinions. For example, the results in *Like* showed a recurring sentiment supporting that the concept of free speech covers all forms of speech, up to and including 'hate speech'. *Dislike* shows several people who disagree with this sentiment, with lines 2, 3, and 9 noting that free speech does not include hate speech and line 7 even suggesting Twitter as an alternative.

Keeping in mind that these particular occurrences have all been heavily downvoted, the results of table 14 also suggests active attempts at limiting what can be said within what the community considers a legitimate 'free speech' context. Seemingly, the notion that 'free speech' covers institutions which have previously been highlighted as illegitimate, such as Antifa, is frowned upon, as exemplified by lines 5 and 6. Furthermore, line 10 shows how a user who is ridiculing a popular conspiracy theory about Hillary Clinton is being heavily downvoted as well.

Concordances of 'hate speech' collocating with 'is' in *Dislike*

		Keyword	
1	un-normal media is where sexist/racist/nazis like u create their oen platform to cry over the fact that NORMAL medias don't allow your	hate speech	.
2	You nazis will **never normalize**	hate speech	. Your oppressive speech will not be tolerated
3	You use	hate speech	towards them **to somehow justify your White supremacy**
4	I plan to exercise my right to free speech by **telling other people who post**	hate speech	**that they shouldn't post it.**
5	There's **a difference between**	hate speech	**and free speech**. If you applied a little brain you might understand that.
6	To Gab owners. Protecting	hate speech	is not commendable.
7	Sick of this	hate speech	you spew. The good news is I've reported your terroristic threats.
8	Why doesn't Gab boot you waste of life, larping Neo Nazis off for	hate speech	already? You morons should try listening to the crap you spit out of your suck hole.
9	Banning	hate speech	isn't censorship. Sorry to burst your altright bubble.
10	You all create your own safe space to spew conspiracy theories and	hate speech	. Nut jobs!

Table 17. Concordances of 'hate speech' collocating with 'is' in *Dislike*

Unlike most previous results, quite a few of the occurrences found in table 17, as well as some in Table 16, present an issue inherent to analysing sentiment based on abstracted delegitimation such as downvotes. When presented with posts which exhibit more than one feature potentially seen as problematic by the Gab community, it is essentially an educated guess as to which sentiment is the cause of this effect. In light of what has been upvoted, in relation to 'hate speech' and 'hate crime' in *Like*, it seems plausible that these downvotes were given on the basis of disagreement with the essential message, i.e. that 'hate speech' is not normal, and that it should be punishable by both Gab and society in a larger sense, as seen in lines 1, 2, 5, 6, 8, 9. At the same time, the above concordances, and indeed the clusters from the overall corpus seen in Table 14, present numerous examples of insults and accusations, which may serve as an alternative reason for negative evaluations. That said, lines such as 4 and 6 arguably do not exude this form of antagonism and have still been downvoted to a significant degree for expressing concerns regarding 'hate speech' and its relation to 'free speech'.

Concordances of 'fake news' in *Dislike*

		Keyword	
1	YEAH! START WITH THE "SHARE FREELY" BULLSHIT ON	FAKE NEWS	#GAB! GONNA TAKE THIS POST TOO, GAB KIKES?
2	OMG there is NO immigrant crime!! Fuck off	FAKE news	@Prisonplanet
3	This is from the Blaze they lied about Iran when Obama was in office stop posting	fake news	about Iran fuck islam fuck Iranian law they get arrested for coed dancing but you idiots are marching us into WW3 wake up.
4	I'm not interested in Drunk Uncle	fake news	YouTube channels.
5	More	fake news	from the lulzcows of the internet /po/ 4chan
6	@MAGA	Fake news	like Breitbart and Infowars. Only republicans would fall for something so stupid
7	@deplorablepatriot it is important to be present on all social media platforms to Correct The Record. There's a lot of	fake news	out there today.
8	Truthfully, this is MSM taking notes from Milo's Breitbart on how to make a shit posting. Trump is shit posting free no matter how many	fake news	reports they make. The Russia connection isn't as it appears!
9	Delete this post. Stop spreading	fake news	. Sweden is beautiful.
10		Fake news	! You extremists are so hell bent on convincing us that Sharia law is coming. Stop the paranoia.

Table 18. Concordances of 'fake news' in *Dislike*

As with some examples in 'hate speech', line 1 of 'fake news' presents yet another problem with analysing sentiments. On the one hand, the post expresses an anti-Semitic sentiment through the derogatory use of the term "kike", which was previously found as a legitimate sentiment in *Like*. On the other hand, this derogatory sentiment is used to negatively categorise Gab as yet another institution controlled by, or at least involving, a Jewish cabal. Given

previous observations, the latter is perhaps more likely, but either interpretation would arguably be shaky.

Obviously, this observation is an outlier, however, with lines 2, 4, 5, 6, 7, 9 and 10 negatively evaluating users on Gab for posting fake news. Particularly, we also see criticism of institutions which have previously been deemed legitimate, such as Breitbart and Infowars, as well as 4chan's /pol/ which has its own popular re-post bot on Gab and individual YouTube channels. Given that all of these were highlighted as extremely frequently cited sources of news on Gab by Lima et al. (2018), this supports their findings that Gab prefers alternative sources of news to mainstream outlets.

6.2 '|x| She'

The occurrence of 'She' brings up an interesting wrinkle to the otherwise easy-to-understand uniformity of the paper so far, as it is not key enough to appear in *Like* but explaining its appearance in *Dislike* is difficult without the context afforded by this reflection. As such, I am going to begin by highlighting the top collocates of both corpora.

Collocates of 'She' ranked by MI in *Dislike*			
MI	**Freq**	**Collocate | Single Word**	**Position**
8.2	14	herself	R
8.2	5	grandma	L
7.4	5	battle	R
7.2	6	mental	R
6.8	6	nasty	R
6.7	95	her	L
6.2	5	beat	R
6.0	5	mention	L
5.9	5	mad	R
5.8	12	die	L
5.5	5	fight	R
5.1	11	lie	R
4.8	8	claim	M
4.1	5	jealous	L
4.0	6	forget	R

Collocates of 'She' ranked by MI in *Like*			
MI	**Freq**	**Collocate | Single Word**	**Position**
8.2	10	mom	M
8.1	5	susan	M
7.4	5	cancer	L
7.3	6	secret	L
6.6	7	die	L
6.5	9	care	L
6.1	6	pray	R
6.0	6	cry	L
5.8	5	mother	R
5.7	7	sell	L
5.5	5	gave	R
5.1	7	sick	R
4.5	11	wife	M
4.3	8	wear	L
4.0	9	care	R

Table 19. Collocates of 'she' in *Like* and *Dislike*

Given the previous mirroring of leftist-progressive ideals in *Dislike* and right-leaning anti-progressivist ones in *Like*, one might very well expect a similar disparity here. To explain this

disparity, I begin by looking at the concordances related to representation, i.e. adjectives such as ‘mental’, ‘nasty’ and ‘jealous’, in *Dislike*.

Concordances of ‘she’ in *Dislike*

		Keyword	
1	She’s nothing but fat	nasty	white trash and she proves it daily!
2	If she thinks those fat	nasty	tits of hers looks good she needs to be in a mental institution
3	Go run and downvote all her lies she spews from that	nasty	evil mouth of ITS.
4	Too bad she don’t keep her	mental	ass on Facebook
5	She really is	mental	. Bat shit crazy. Maybe she really thinks she is all those people
6	She really has	mental	issues.
7	She	jealous	cuz she can never have kids
8	Tell the truth I AM	JEALOUS	OF TRACIE AND EVERYTHING THAT SHE HAS
9	Not to mention what a creepy	jealous	fan girl she is.

Table 20. Concordances of 'she' in *Dislike*

While it is extremely difficult to draw any broader conclusions from this particular observation, what appears to be happening in *Dislike* is people attacking women with some clout on the platform. It is possible, for example, that these are women who follow the common social practices on Gab and are therefore deemed defensible, but it is honestly impossible to tell without having access to the full context of the various posts. What is interesting, however, is the overall negative prosody implied by the collocates in *Dislike*, especially compared to the overall neutral and even positive prosody in *Like*, with verbs such as ‘care’ and ‘pray’. At closer inspection, however, these positive appraisals appear to relate to a particular incident, in the same way as some earlier observations, regarding a female member of Gab who died of breast cancer.

While the brief examination of the female pronoun ‘she’ highlights an interesting discrepancy between legitimate and illegitimate representation of females on Gab, it is arguably speculative and highly flawed. This is mostly due to the nature of the sub-corpora. Further studies analysing sex and gender on Gab would likely be better off analysing the broader corpus, rather than delegitimised speech in particular.

6.2.1 *Controversial*

This last corpus is tiny, at only 600 post and 5,000 words. The relatively tiny size of the *Controversial* sub-corpus means that the repeating the approach taken with *Like* and *Dislike* does not make sense, given that all but four words have frequencies lower than three. Instead of searching for, and analysing, keywords, what follows was extracted by extracting 25 lines of

text which had a sum total of between -20 and +20 upvotes when subtracting downvotes from upvotes. This theoretically leaves the sentiments most closely contested by the community, while still requiring at least 60 people to have made an evaluative contribution through either downvoting or upvoting. The analysis of *Controversial*, then, follows a more qualitative approach to analysis.

Extracted Concordances from *Controversial* +20 to -20

#	↑	↓	**Body**
1	43	43	Ever notice how far right guys are just like far left guys- Both use SJW tactics and try to SHAME you into believing their BS - BLM and WHITE GENOCIDE different sides of the same coin - There are good black guys, white guys, and Jews- mind your own biz- Racism is an mental illness- Just sayin' #MAGA
2	33	33	I told you to boycott all Google sites. They're so SJW it pains them to even acknowledge CHRISTMAS. Their loathsome 'doodles' honor every phony baloney woman scientist. They are the total enemy of the right.
3	33	33	I'm simply telling men not to choose a wife from the whore pool. Men have to stop wanting to test the goods if they want to marry a good girl. One of my friends (attractive, btw) didn't marry until her 30s because she'd eventually be rejected because she wouldn't put out.
4	48	47	Leftists have equally crazy factions show up to all major rallies. It doesn't hurt them because they have their own media - that is, the 'mainstream' - on their side. When we have our own media we can fail to show our wackos too.
5	44	41	It is happening because you're an utter jackass that no one wants to defend, who claims that anyone who is dumb enough to defend you is a secret Nazi. Most people would rather leave the Internet to the SJWs than stand with a Fake Right Clown Nazi like you.
6	63	56	Hi All, we've just delisted a video that called for the immediate extermination of Jewish people in a "race war now". via @Azzmador This video was also trending for the day. Please note our guidelines state that: Threatening, inciting, promoting or glorifying violence will not be tolerated.
7	52	43	I think he is a Fake Right Clown Nazi. I don't know what he truly believes, but I doubt he actually knows much about what National Socialism genuinely is.
8	77	57	The only fraud is "Andrew Anglin". I think @a should censor anything he wants. And I think he should go ahead and ban Andrew Anglin for attempting to destroy Gab. Why do you guys pay those losers any attention at all?
9	50	32	So much this- we have got to change the laws and marriage needs to become a contract once again. Break the contract and you're heavily penalized. Make it hurt to get divorced. There is very little a couple can't work out and walking away shouldn't be easy or financially beneficial.
10	138	132	Why does the establishment media keep giving white supremacists a platform?
11	41	31	Another reason why monogamy and marriage are beneficial and necessary for a working society
12	39	59	"Weev was banned months and months before The only people butt-hurt are shills like you, whinging every week for the return of your cult leader

			News for you: nobody gives a fuck The world is bigger than Weev"
13	32	43	LMFAO. Looks like we have a whole bunch of snowflakes here on Gab. Look how triggered the anti-semites became once they were called out for the nasty people they are.

Unsurprisingly, the discourse found in *Controversial* is highly divided, with an almost even split mirroring both *Like* and *Dislike*. Encouragingly, it reflects a degree of intolerance towards blatant sexism and racism, at least on some occasions, as exemplified by lines 1 and 3. Arguably, however, since the corpus is tiny, it is quite likely that these are exceptions to the rule rather than examples of the rule itself. Conversely, a highly recurring phenomenon in this sub-corpus is that of pro-right people seeking some form of middle ground between the new right or alt-right. and people who identify as Nazis. Examples of this can be seen in lines 6, 7, 8, 10, 12 and 13. This form of right-wing centrism, for lack of a better term, is nowhere to be found in either of the previously examined sub-corpora, and highlights seemingly divided opinions on whether anti-Semitism and open white nationalism should be allowed on Gab. The discourse observed here all centres around a particular event in time; the banning of Gab user @weev, also known as Andrew Anglin, on the grounds of promoting violent and terrorist ideas, following threats by Gab's then registrar Asia Registry that they would refuse to host the site until the content posted by Anglin was removed. Anglin, who openly wears the mantly of anti-Semite and neo-nazi, is well known within white nationalist circles as the creator and editor of The Daily Stormer, a website which is classified on MediaBiasFact as an 'extreme right hate group' (Media Bias Ratings, 2018 "Daily Stormer"). It is worth noting that the discussion surrounding this issue goes on for far longer than the controversy itself, with user posts referencing the banning of Anglin running September 19th 2017, the day of the banning, to at least the beginning of 2018. This shows the continued divide created by this decision, likely because of what it conveys regarding Gab's own definition of 'free speech'.

7. Discussion

The following section discusses the findings in relation to the two stated goals of the present study, beginning with discourse and legitimation and ending with a brief examination of the utilisation of Gab as a corpus of alt-right discourse.

Table 21. Concordances extracted from *Controversial*

7.1 Discourse and Legitimation

The primary function of the present study has been to examine the discursive representations that make up the structure of the community found on Gab. In this regard, the analysis highlights several interesting findings which mostly appear to support the results found in previous studies by Lima et al. (2018) and Zannettou et al. (2018). While I hold that their approach may be flawed, in the sense that they ignore the context of the words from which they draw their analysis, the results of Zannettou et al. in particular seem coherent with those found in the present study.

The keyword clusters found in *Like* clearly indicate numerous discourses which can be considered synonymous with the growing alt-lite, such as a rejection of mainstream institutions, support for unlimited free speech and the rejection of 'hate speech' and 'hate crime' as valid terminology. According to the most recurrently upvoted comments on Gab, the mere definition of 'hate speech' is problematic and throughout *Like*, several people construe the use of the 'hate speech' label tool used to stifle 'free speech' by disallowing certain discourses which would otherwise naturally flourish. Beyond that, however, the recurrent use of 'white genocide' and 'deep state', as well as the attitude rendered towards these subjects, suggests that a majority of users seem to further support ideas which are in accordance with alt-right and white supremacist tactics. The positive connotations associated with 'Pizzagate' and the rejection of institutions which claim that this topic is 'fake news' also suggests that the community is prone to supporting dangerous conspiracy theories, which may ultimately foster further violent behaviour.

Importantly, the findings also suggest that Gab functions as what has colloquially been termed a "filter-bubble", largely as a result of recurrent downvoting which appears to be a conscious (de)legitimation strategy. While some might argue that downvoting is an individual practice, rather than a social one, as delineated in section 3.2, there are a few observations which support this conclusion of social practice, beyond what is highlighted in the analysis:

First, downvoting itself entails yet another interesting technical difference between Gab and Twitter in that it affects not just the score of the post but also the score of the user. On the one hand, this means that if a user is sufficiently downvoted, they are unable, themselves, to render any form of downvoting of other people. On the other, it means that downvoting other people requires the user to sacrifice a portion of their own score in order to actively express disagreement of other users. This second part leads to an interesting but un-explored keyword in *Like*; #upvotebomb, which was found to be leveraged by people on occasions where they

might be losing too much score through either downvoting or being downvoted. Further analysis of the user-networks may highlight whether this presents a systematic attempt at silencing certain opinions on behalf of a particular group of users.

From the observations made throughout the analysis, however, it seems plausible to conclude that the act of downvoting and upvoting on Gab mirrors the ideological make-up of the site in a broader sense. Were this not the case, one would expect to find some occurrences of oppositional or at least neutral topics and attitudes within *Like*. Granted, a small but significant portion of the actively delegitimised content can be considered spam or attempts at trolling, i.e. those containing insults and repeated emoji-use. That said, however, the topics and attitudes expressed in *Dislike* show that this exact opposition to popular sentiments is systematically silenced by the existing majority on the site. In opposition to *Like, Dislike* contains the majority of people who disagree with the notion of unlimited 'free speech' and who reject the notion that 'hate speech' belongs under the first amendment. The institutions which are seen as legitimate by those who have been downvoted are also often the same as those deemed illegitimate in *Like,* including Antifa and CNN. Ultimately, this indicates that even if Gab itself may adhere to the common understanding of 'free speech', the implementation of their user-moderation system fails to account for community imbalance. Consequently, topics and attitudes which are deemed illegitimate by a large enough majority may be systematically silenced until they disappear completely. In this sense, downvoting appears to serve a systemic purpose, since the process takes away the target's user-score, in turn preventing the targeted user from downvoting posts that they may disagree with. This mechanic ultimately leads to a filter-bubble where the community majority sets the limits of what is considered acceptable speech. The collocations with 'free speech' in *Dislike* highlights that there is no room on Gab for people with what may be considered a soft belief in 'free speech', and that users who express concerns regarding 'hate speech' will find themselves without a voice on the site. Unsurprisingly, then, the use of downvoting occurs even on occasions where the posts show no obvious use of insults or derogatory language, such as when a user proclaims, "I plan to exercise my right to free speech by telling other people who post hate speech that they shouldn't post it." (Table 17, line 4). This further supports the theory that it is the semantic content, i.e. the opinion of the person, which is found disagreeable, rather than the presentation of the opinion itself, as may very well be the case with the posts tied to derogatory language such as 'evasive chickenshit' and 'nazi trash'.

As a last side-note, it seems somewhat interesting that very little of the recurrent discourse observed throughout the analysis has been overtly hateful. Barring a few individual

examples of posts using the word ‘kike’ and the ableist use of the word ‘retarded’ the discourse has mostly lacked what most would call overt hate speech. Instead, the discourse highlights occurrences of far less overt language which still arguably suggest similar convictions, such as the recurring use of ‘deep state’ to suggest Jewish involvement and Table 10, line 4 noting that “Jewish values are incompatible with American values”. While this does not mean that overt hate speech does not exist on Gab – ongoing studies show that it most assuredly does – this may imply that such overt hatred does not naturally float to the top. Instead, it is likely that this form of discourse is kept in smaller, even tighter-knit sub-communities on the site.

7.2 A Corpus of the Alt-Right

As alluded to numerous times throughout the paper, the secondary goal of this examination has been to ascertain whether Gab serves as a valid foundation for a corpus of alt-right communication and sentiments. While the observations made above would indicate that this is indeed the case, the findings made in both *Dislike* and *Controversial*, while deemed illegitimate by the broader community, does reflect some form of counter-discourse present on Gab. Even if counter-discursive sentiments are a minority, then, this still presents a potential challenge with wholesale observation of Gab as an alt-right community. That said, the data-isolation approach used to create the three sub-corpora used in the present study would possibly allow for the isolation of such counter-discourse, given the relative homogeneity of the site users.

8. Conclusion

The present study set out to examine the website Gab.ai in order to answer the following two thesis questions:

1. Which discourses are typified and legitimised, as opposed to rejected or delegitimised, by the Gab community, and in what ways do the findings support or reject previous observations and accusations of alt-right homogeneity on the site?
2. What obstacles, if any, hinder the utilisation of Gab as the foundation for a corpus of alt-right social media interactions?

To do so, the study presented an expanded version of the corpus used in Zannettou et al. (2018) and analysed the most liked, disliked and controversial sentiments present on the platform. The results highlighted by the analysis present a picture of a largely homogenous alt-right community, mostly supporting the characterisations rendered by both Lima et al. (2018)

and Zannettou et al. (2018). Furthermore, this conclusion lends validity to the use of Gab as the foundation of an alt-right social media corpus.

As expected, some of the most discussed topics on Gab are the concepts surrounding 'free speech' and 'hate speech', mirroring the focus of the website itself. But whereas the creator of Gab purports to support the concept of dialogical free speech, and claims to welcome all who wishes to join, this openness to dialogue is not reflected on the site itself. Instead, who support an unlimited form of free speech, which includes hateful and pejorative speech, find themselves upvoted, whilst the ones who disagree with this notion are found to have been downvoted. From there, this mirrored (de)legitimation strategy permeates all of the analysis, creating a view of a community which utilises the tools afforded to them, namely upvoting and downvoting, to delimit the type of speech allowed on the site. This, in combination with recurrent alt-right sentiments, such as 'civil war', 'deep state', 'red pill' and 'white genocide', in the more frequent clusters serves to solidify the conclusion that the platform has an established and homogenous view of the world. While it is difficult to say with certainty whether this political homogeneity is more 'alt-right' or 'alt-lite', the characteristics highlighted in the present study largely mirrors the definition rendered of the alt-right (Anti-Defamation League, "Naming the hate").

Despite the tentative nature of the study, the results support further investigations into Gab as an alt-right platform, especially from a discursive point of view. Although Twitter and Facebook remain better platforms for viewing the interaction between opposing opinions, Gab likely serves as a uniquely popular outlet with a largely homogenous community.

9. Issues and Future Studies

Even in the face of the relatively narrow scope of the present study, there are several issues with the observations made herein. It should be noted, of course, that these issues arguably extend to any observations made of social media at the present, but that does not make them any less noteworthy. The following section shows some of the flaws inherent to the approach taken in the present study, particularly in relation to the conclusions made above, and, where possible, amendments through further studies are suggested.

9.1 A note on Bots

While the study has dealt with a number of issues inherent to social media texts and corpora in general, one issue has been too pervasive to be dealt with here: the unknown constant of social bots. Anecdotally, the account set up on Gab entirely to facilitate the scraping process

of this paper, ended up with more than a dozen followers, having never posted a single post. Most of these were adult-themed bots presumably programmed to follow every user on the site as an attempt at phishing. While these bots follow easily identifiable patterns, though, the same cannot be said to be the case for malicious social bots. The issue of social bots concerns more than simply Gab, of course, with for example Pew noting that two-thirds of all shared links on Twitter are by bots (Wojcik et al. 2018). What does concern Gab in particular, however, is the fact that bots have been found to be generally aimed towards increasing exposure to negative and inflammatory content (Stella, Ferrara and Domenico 2018). As was highlighted in the various parts of the analysis, a large part of the discourse on Gab is contentious and controversial. Since these are key characteristics of the work done by the Russian Internet Research Agency (IRA) and, according to Stella, Ferrara and Domenico, a multitude of general-use bots, it is more than likely true that some portion of the discourse observed in the present study would be the result of automated programs rather than human beings. As such, bots must be considered in this evaluation as well, but the process of doing so is simply beyond what is feasible within the allotted timeframe of the present study.

To this end, a future study has to consider the differences in discourse patterns between bots and humans, perhaps inspired by the data released by Twitter (Twitter, 2018) and possibly even attempt to identify and isolate bot behaviour. Following identification, a future study may also benefit from analysing whether humans adopt the discourse sown by bots and legitimise it further, through sharing, replying or upvoting.

9.2 A Note on Multimodality

While the data utilised in the present study is a substantive and comprehensive sample of the written discourse present on Gab, it, as most other studies of social media, fails to represent any portion of the multimodal aspect of social media. On both Gab and Twitter, a significant portion of the posts made by users consists of either images or videos in combination with text, or – as is often the case – images or videos exclusively, with no contextual representation through text. While analysing images qualitatively is possible, through the website itself, this would give no broader understanding of the way images are used to portray legitimate or illegitimate representations of the written discourse highlighted in the present study. As a result, the present study fails to account for a significant portion of non-verbal discourse which may be paramount to understanding Gab.

At present, no widely accepted solution exists to this problem, but early forays into the use of vision-based image-corpora by Christiansen and Dance suggest that this may very

well be possible in the future. As such, future studies should be encouraged to consider the importance of recurrent image-based representation, especially through memetic imagery.

Works Cited

Alexa. (2018, May 21). *gab.ai Traffic Statistics.* Retrieved from Alexa: http://alexa.com

Anti-Defamation League. (2018, February). *Anti-Semitic Incidents Surged Nearly 60% in 2017, According to New ADL Report.* Retrieved from ADL Homepage: http://www.adl.org

Anti-Defamation League. (n.d.). *Echo.* Retrieved from Anti-Defamation League: https://www.adl.org

Baker, P. (2006). *Using Corpora in Discourse Analysis.* London: Continuum.

Baker, P., & McEnery, T. (2015). *Corpora and Discourse: Integrating Discourse and Corpora.* London: Palgrave.

Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Berger, J. M. (2016). *Nazis vs. ISIS on Twitter: A Comparative Study of White Nationalists and ISIS Online Social Media Networks.* Washington D.C.: George Washington University: Program on Extremism.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Brandom, R. (2018, August 9). *Microsoft threatened to drop hosting for Gab over hate speech posts.* Retrieved from The Verge: http://www.theverge.com

Bray, H. (2016, September 14). *For those tired of online censors, an (alt-right) alternative.* Retrieved from The Boston Globe: http://www.bostonglobe.com

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 139-173.

Bryden, J., Funk, S., & Jansen, V. A. (2013). Word usage mirrors community structure in the online social network Twitter. *EPJ Data Science*.

Chetty, N., & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and Violent Behaviour*, 108-118.

Christiansen, A. (Unpublished). #FakeNews: Analysing 'fake news'-in-use on Twitter. *Unpublished*.

Collins, M. (2018, March 2). *The collapse of the BNP and EDL has made the far right deadlier.* Retrieved from The Guardian: http://www.theguardian.com

Dearden, L. (2016, November 22). *White supremacists chant 'hail Trump' while performing Hitler salutes at alt-right conference.* Retrieved from The Independent: http://www.independent.co.uk

Ellis, E. G. (2016, September 14). *Gab, The Alt-Right's Very Own Twitter, Is The Ultimate Filter Bubble.* Retrieved from Wired: http://www.wired.com

Fairclough, N. (2012). *Critical Discourse Analysis.* Retrieved from Academia: http://ww.academia.edu

Fairclough, N. (2015). *Dialectical Reasoning in Critical Social Analysis and Critical Discourse Analysis.* Retrieved from Academia: http://www.academia.edu

Fairclough, N. (2017). *CDA as dialectical reasoning: critique, explanation and action.* Retrieved from Academia: http://www.academia.edu

Fioretti, J. (2018, January 19). *Social media companies accelerate removals of online hate speech -EU.* Retrieved from Reuters: http://www.reuters.com

From Alt Right to Alt Lite: Naming the Hate. (n.d.). Retrieved from Anti-Defamation League: http://www.adl.org

Goodwin, M. (2011, November 2). *Far-right extremism is much more than a political irritant.* Retrieved from The Guardian: http://www.theguardian.com

Gray, P. W. (2018). 'The fire rises': identity, the alt-right and intersectionality. *Journal of Political Ideologies*, 141-156.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: from user discussions to theoretical concepts. *Journal of Politeness Research*, 215-242.

Hardaker, C., & McGlashan, M. (2016). "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics*, 80-93.

Hern, A. (2017, July 12). *Extremists driven off Facebook and Twitter targeting smaller firms.* Retrieved from The Guardian: http://www.theguardian.com

Holden, M. (2018, February 26). *Britain is facing serious far-right terrorism threat, says top UK officer.* Retrieved from Reuters UK: https://uk.reuters.com

Johnson, G. (2016, November). *The Alt Right: Obituary for a Brand?* Retrieved from Counter-Currents Publishing: http://www.counter-currents.com

Johnson, G. (2017, January). *White Nationalism, The Alt Right & the Alt Lite.* Retrieved from Counter-Currents Publishings: http://www.counter-currents.com

Kantrowitz, A. (2016, September 6). *This New Social Network Promises Almost-Total Free Speech To Its Users.* Retrieved from BuzzFeed News: http://www.buzzfeednews.com

Keyser, Z. (2018, October 28). *Gab, The 'Alt-Right Twitter' Used By Antisemitic Pittsburgh Shooter.* Retrieved from The Jerusalem Post: http://www.jpost.com

Kotenko, J. (2013, May 10). *According to a new report, Twitter is a breeding ground for terrorism and hate speech.* Retrieved from Digital Trends: http://www.digitaltrends.com

Lapowsky, I., & Levy, S. (2018, April 24). *Here's What Facebook Won't Let You Post.* Retrieved from Wired: http://www.wired.com

Lee, T. B. (2017, August 18). *Google explains why it banned the app for Gab, a right-wing Twitter rival.* Retrieved from ArsTechnica: http://www.arstechnica.com

Lehmberg, T., & Wörner, K. (2008). Annotation Standards. In A. Lüdeling, & M. Kytö, *Corpus Linguistics: An International Handbook* (pp. 484-500). Berlin: Walter de Gruyter.

Lima, L., Reis, J. C., Melo, P., Mural, F., Araujo, L., Vikatos, P., & Benevenuto, F. (2018). Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 515-522). IEEE.

McEnery, T., & McGlashan, A. (2013). *Corpus: Some Key Terms.* Lancaster: Centre for Corpus Approaches to Social Sciences.

McEnery, T., McGlashan, M., & Love, R. (2015). Press and social media reaction to ideologically inspired murder: the case of Lee Rigby. *Discourse & Communication*, 237-259.

Price, R. (2017, September 8). *The chilling resurgence of Britain's Far Right: UK extremists are on the march again - and they're recruiting middle-class female graduates to push their warped ideology.* Retrieved from Daily Mail: http://www.dailymail.co.uk

Rawlinson, K. (2018, March 7). *Britain First leaders jailed over anti-Muslim hate crimes.* Retrieved from The Guardian: http://www.theguardian.com

Sharman, J. (2018, March 28). *Tommy Robinson permanently banned from Twitter.* Retrieved from Independent: http://www.independent.co.uk

Stella, M., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences of the United States of America 2018*, (pp. 12435-12440). New York.

The British Association for Applied Linguistics (BAAL). (2006). *Recommendations on Good Practice in Applied Linguistics.* Retrieved from BAALweb: https://www.baalweb.files.wordpress.com

The Home of Free Speech Online. (2018, October). Retrieved from StartEngine: http://www.startengine.com

Toor, A. (2017, August 18). *Google removes Gab app for violating hate speech policy.* Retrieved from The Verge: http://www.theverge.com

Torba, A. (2016, November 30). *Our Interview With Fake News Publication The New York Times.* Retrieved from Medium: http://www.medium.com

Trampuš, M., & Novak, B. (2012). Internals of aggregated web news feed. *Proceedings of the Fifteenth International Information Science Conference IS SiKDD 2012*, (pp. 431-434).

Twitter. (2017, January). *Government TOS reports.* Retrieved from Twitter: http://transparency.twitter.com

Twitter. (2018, October). *Elections integrity: Twitter's focus is on a healthy public conversation.* Retrieved from Twitter - About: https://about.twitter.com

Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis.* Oxford: Oxford University Press.

Vathi, E., Siolas, G., & Stafylopatis, A. (2017). Mining and categorising interesting topics in Twitter communities. *Journal of Intelligent & Fuzzy Systems,* 1265-1275.

Ward, C. (2018, November 28). *Anti-Semitism never disappeared in Europe. It's alive and kicking.* Retrieved from CNN: https://edition.cnn.com

Yates, W. (2016, December 14). *Gab: Free speech haven or alt-right safe space?* Retrieved from BBC Trending: http://www.bbc.co.uk

Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringhini, G., & Blackburn, J. (2018). What is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber? *International World Wide Web Conference 2018* (pp. 1007-1014). Lyon, France: International World Wide Web Conferences Steering Committee.