Investigating the Far-Right Online: Using Text Data to Understand Online Subcultures

This text is chapter five of ten in the publication *Investigative Methods: An NCRM Innovation Collection*.

How to cite this document

Brace, L. (2022) Investigating the far-right online: Using text data to understand online subcultures. In: Mair, M., Meckin, R. and Elliot, M. (Eds) *Investigative Methods: An NCRM Innovation Collection*. Southampton: National Centre for Research Methods, pp. 54-65. DOI: 10.5258/NCRM/NCRM.00004548

5. Investigating the Far-Right Online: Using Text Data to Understand Online Subcultures

Lewys Brace (University of Exeter)

This contribution provides an introduction for social science researchers on the use of computational methods within investigative research for analysing large text corpora to develop an understanding of online communities and subcultures. It offers a case study of the MineChans project, which utilised such methods in investigating the relationship between the online discussions on a collection of anonymous image-board forums, including 4chan and 8chan, and real-world, offline, attacks by right-wing extremists, making these forums a radicalising milieu. While these analytical techniques are new, they are actually fairly easy for social researchers to implement due to the nature of contemporary high-level programming languages such as Python.

Introduction

It has long since been established within the social sciences that subcultures have their own distinct norms, values, traditions, and even languages that differentiate them from the dominant culture, and that membership of a subculture can influence an individual's behaviour as they adapt their beliefs and attitudes to better match that of a given subculture (Herbert 1998; Haenfler 2013). The exponential increase in both the development and use of the internet and digital technologies during the last 20 years has enabled the emergence of a range of distinctive online subcultures tailored to specific interests. Communication technologies also allow members of such subcultures to communicate anonymously, thus granting them opportunities to engage in behaviours and share subcultural knowledge without the threat of either rejection or detection (Blevins and Holt 2009; Holt and Copes 2010). An online subculture that has made use of these aspects of digital technologies is the contemporary far-right (Baele et al. 2020b), and this has not only resulted in changes to the online-offline interplay of ideas and actions, with Europol noting that "online spaces have been observed to strengthen international links among right-wing extremists" (2020, 68), but has also allowed for the rapid growth and diversification of digital platforms by which these notions can spread; with transnational relationships between content producers (Davey and Ebner 2017) and the livestreaming of attacks by right-wing extremists (RWE) on "grey" chatrooms (Evans 2019b; Baele et al. 2020a, 2020b, 2021; Brace 2021) being just two examples.

Our lack of understanding of the relationship between the behaviour of the far-right online and RWE attacks underlines the importance of developing investigative studies exploring the behavioural dynamics of online subcultures. However, this requires researchers to conduct large-scale, systematic analysis of different aspects of online communities, particularly the nature of their discussions, in order to ascertain how these online interactions can give rise to an online radicalising milieu. The aim of this paper is to show how social researchers can use methods from computer science to analyse large text corpora as part of investigations of that kind. While a number of social researchers already use such methods, this paper aims to demonstrate to those who are unfamiliar with these methods how developing even a basic understanding of a high-level interpretative programming language, specifically Python, can enable them to carry out complex, often cutting-edge, analysis of large text corpora with relative ease. This is a particularly timely consideration, as researchers adapt to deal with "big data," which in the context of social science, often refers to large textual datasets; data that has been extracted from Twitter or YouTube comments, for example. This discussion will be embedded within a case study concerning far-right online extremism, and will demonstrate how social researchers can utilise Python to conduct the kind of research into online communities and subcultures which previously would have been incredibly time consuming and fraught with challenges.

Background

At 13:28 (NZDT) on 15th March 2019, a user posted a message to the 8chan/pol board stating "Well lads, it's time to stop shitposting and time to make a real life effort post" before going on to declare that "I will carry out and [*sic*] attack against the invaders." Moments later, Brenton Tarrant began his attack on two mosques in Christchurch, New Zealand, which resulted in the deaths of 51 people. It is known that Tarrant was the one who made this post to 8chan/pol because it included links to his manifesto and a Facebook live stream of his attack. Tarrant's manifesto itself contained extensive references to the 8chan/pol subculture and its associated in-group/out-group terms and "in-jokes," a clear sign that he had been an active participant on the forum and engaged with its subcultural practices (Evans 2019b; Baele et al. 2020a). Unfortunately, this *modus operandi* of posting a message to 8chan/pol announcing an individual's intentions to carry out an imminent attack and their justification for doing so would be repeated by other individuals responsible for right-wing extremist (RWE) acts of violence. These include the Poway Synagogue (27/04/2019) and El Paso Walmart (03/08/2019) shooters.

There are three aspects of the Christchurch case that demonstrate the way in which digital platforms, such as 8chan/pol, are no longer places of "harmless" and "edgy" online conversations, but instead constitute a radicalising milieu through their subcultural dynamics. First, Tarrant's terrorist activity was inspired by discussions on such platforms. Second, he used this online forum to advertise his ideas and actions. Third, he has subsequently become known as a "hero" or "saint" on these forums and has seemingly inspired others to adopt his *modus operandi* (Baele et al. 2020a).

It was within this context that 8chan acquired notoriety in the public domain and came to the attention of security and law enforcement practitioners. The site itself was one of a family of sites that also include 4chan, 8kun (8chan's direct successor once it was shut down), 16chan, Endchan, Neinchan, etc. These are colloquially referred to as "the chans," and are an ever-changing and expanding group of sites, all of which are anonymous image-boards with a near identical visual layout (Baele et al. 2021). The anonymous aspect of the chans stems from the fact that users of these sites do not have usernames, and any posts they make are therefore anonymous. This is a trait of the forums that users take very seriously, with the hacktivist group known as "Anonymous," which originated on 4chan, drawing its name from it. The image-board aspect of the chans then refers to the way in which the posts made to these sites often contain both text and images, and sometimes only images.

Each chan iteration is made up of multiple boards dedicated to different topics, including video games, movies, cooking, pornography, etc. In the context of RWE, the boards that are of most interest to security and law enforcement practitioners are the /pol (for "politically incorrect") boards. Nearly all chan iterations have their own version of the /pol board, which are ostensibly online forums where users can discuss matters pertaining to society and politics in an environment that they claim is "free from political correctness," but this is little more than a euphemism for racist, misogynistic, and anti-Semitic discussions (Baele et al. 2021; Brace 2021).

The 8chan/pol iteration of these boards grew in popularity following the "#gamergate" controversy when 4chan moderators decided to ban conversations on the subject. This triggered a migration of 4chan users to 8chan, with the latter marketing itself as a "no holds barred" equivalent to the former. This lack of moderation and anonymity on 8chan quickly resulted in its /pol board becoming a hub for RWE (Conway et al. 2019). The 8chan iteration was shut down on 5th August 2019, two days after the El Paso Walmart shooting. However, this was accompanied by an increase and diversification in the number of new chan iterations coming online, all of which had boards containing RWE content.

The MineChans Project

Some useful insights into the nature of the chan /pol boards and their subculture were gained from both early academic work using qualitative analyses of a small fragment of the boards' content (Trammell 2014; Massanari 2017; Nagle 2017; Ludemann 2018; Merrin 2019) and from investigative research using various open-source intelligence techniques, one example being the specialised news outlet Bellingcat (Evans 2019a, 2019b, 2019c). However, the lack of a large-scale, systematic analysis of the content of these forums left a number of unanswered academic questions that would be of interest to those researching the online far-right; such as what narratives and themes are present in the content of these boards and what is the nature of the relationship between online discussions, the discursive and potentially radicalising milieu that emerges from them, and real-world, offline, attacks?

The MineChans project[1] sought to investigate such questions with the dual aim of advancing our academic understanding of the nature of the online far-right and developing both general theories of radicalisation (Horgan 2014; McCauley and Moskalenko 2017; Kruglanski et al. 2019) and online radicalisation in particular (Conway 2016; Fernandez et al. 2018; Reeve 2019). There was also the intention to use methods that would allow law enforcement and security practitioners to fine-tune their analysis of these forums and tailor their interventions more effectively, in terms of measures such as implementing counter-narratives and shutting down specific sites.

The MineChans team used several computer-assisted analysis techniques to analyse the full text corpus of different chan /pol board iterations. This first involved extracting the full text and image data from 4chan/pol, 8chan/pol, 8kun/pol, 16chan/pol, Infinitychan/bestpol, Endchan/pol, and Neinchan/pol. These specific iterations were selected because, at the time, they had been identified as the most important sites by a range of government and law enforcement units[2] and due to them being the most active in terms of average number of daily posts at the time of data collection.

This data collection process generated a large amount of text data that needed to be analysed if the MineChans team were to understand the posting behaviours and the underlying subculture of these forums. Having to analyse large amounts of text data such as this is a problem that social scientists are increasingly being faced with when investigating online communities, and even more so in the age of "big data"; a term that for those investigating online social phenomena often refers to large amounts of text data obtained from sources such as Twitter. Indeed, the ever-increasing prominence of the internet and digital platforms in our everyday lives offers new possibilities and avenues of research for exploring many contemporary societal issues. While human analysts are able to read such text data and interpret it within specific socio-cultural contexts, the vast amount of data generated by our online behaviours means that social investigators now require new computational methods to analyse such textual data and unearth patterns at the aggregate level (Nguyen et al. 2020). This need has resulted in a boom in the adoption of methods from computer science by social scientists in recent years (O'Connor et al. 2011; Zhang et al. 2020). While it is worth mentioning that small differences between social scientists and computer scientists exist in regards to differences in perspective, these differences do not constitute a disciplinary divide (DiMaggio 2015), and the adoption of these methods by the social sciences has been incredibly successful in terms of productive research, as well as giving rise to the sub-discipline of Computational Social Science (CSS) (Edelmann et al. 2020; Lazer et al. 2020; Nguyen et al. 2020; Zhang et al. 2020).

However, while these advances in CSS have been impressive, there will inevitably be some social scientists who wish to investigate phenomenon such as online subcultures, but who do not (yet) possess the relevant skills to do so. While specialised software exists that allows for analysis of text data, i.e. NVivo and Atlas.ti, most of these options are closed-source, meaning you have to pay for them, and these options often lack the ability and computational power to implement many cutting-edge analytical techniques. In contrast, there are a number of high-level programming languages which have specialised modules that allow for large-scale textual analysis *and* are open-source; meaning they are free to use. Currently, the most popular high-level programming language amongst data scientists and academic researchers is Python, which gets its name from Monty Python's Flying Circus. Python's popularity is due to a number of its features, with the one most pertinent to our discussions here being an emphasis on importing modules. Modules are specific pieces of code that have been developed by users with a specific task in mind. This means that users can often import a module into their code that allows them to carry out specialised operations instead of having to code everything "from the bottom up." An example of a module is the tweepy module, which allows users to quickly develop a web scraper that interacts with the Twitter API[3] to collect Tweets and Twitter user data.

Using Python to Analyse 8chan/pol

Using Python, the MineChans team developed a custom web-scraper that was used to collect visual and textual data from posts, along with each of the posts' metadata (time and date of post, etc.). This custom built scraper used a mixture of Python modules, specifically the request module and the BeautifulSoup module. Any reader who is interested in seeing how simple it is to construct a web scraper in Python using these modules can download a copy of the scraper used in the work discussed here from the author's GitHub.

This scraper was used to extract the entirety of the 8chan/pol text corpus, from the time that this iteration first came online on 19th April 2017 until it was shut down on 5th August 2019. Data was collected for all original posts and reply posts, resulting in 435,697 total posts being collected, including all the corresponding metadata for each post (key information such as its date and time, etc.).

This 8chan/pol data was then analysed to allow the researchers to construct a "birds-eye view" of the entire text corpus, develop an understanding of the /pol subculture, and to then examine the impact of the Christchurch shooting on the forums' content and pace. What follows is intended to serve as a demonstration of how textual data analysis can aid social researchers in developing an understanding of an online subculture by exploring a forum's content and the dynamics of its online community, and how this can be done with relative ease using computational methods implemented in a high-level programming language such as Python. The interested reader can find the full analysis of this forum in Baele et al. (2020a).

In measuring the popularity of the forum, the number of original posts made per day was used as a proxy measure for how "busy" the forum was. This was done using the metadata of posts, with timestamps being harmonised into the UTC time zone (as this is the time zone used by the forum itself). The researchers then plotted the number of posts across time to show variation during the whole lifespan of the website. Although, the anonymous nature of the forum could mean that a single or small number of individuals making numerous posts per day could have made the forum look more popular than it was, this is the best that could be done given the anonymity of the userbase. However, this simple piece of analysis proved sufficient in allowing the researchers to begin assessing general questions on posting dynamics.

Indeed, as Figure 1 shows, this part of the analysis demonstrated that it took a couple of months for the board to gain traction and start receiving a steady number of posts per day during the summer of 2017, with the pace of the forum increasing to around 30 new original posts per day (not counting replies) after that. The graph shows a significant surge in activity lasting for a couple of days beginning on 27th May 2018, which corresponds to the first major discussion of the far-right "QAnon" conspiracy theory. This is the belief that the world is controlled by a global cabal that is anti-Trump and which operate an international child sex-trafficking ring. While Donald Trump was president, the theory stated that he was planning to dismantle the cabal through a series of simultaneous arrests that QAnon followers refer to as "the storm." This conspiracy theory has been undergoing a transformation since Trump's defeat in the presidential election in November 2020, with the end point of this transformation still to be determined. The two other substantial increases in pace then occur on 15th March 2019, the day of the Christchurch shooting, and a more gradual increase from mid-June to mid-July 2019. This is followed by a drop in original posts shortly before the El Paso Walmart attack in August 2019.

This simple piece of data manipulation and visualisation enabled the MineChans team to demonstrate that not all attacks had the same impact on the forum's pace and dynamics. Indeed, only the Christchurch attack had a clear impact in this regard. It also demonstrated that 8chan/pol's popularity was not significantly affected by the shootings (Baele et al. 2020a).

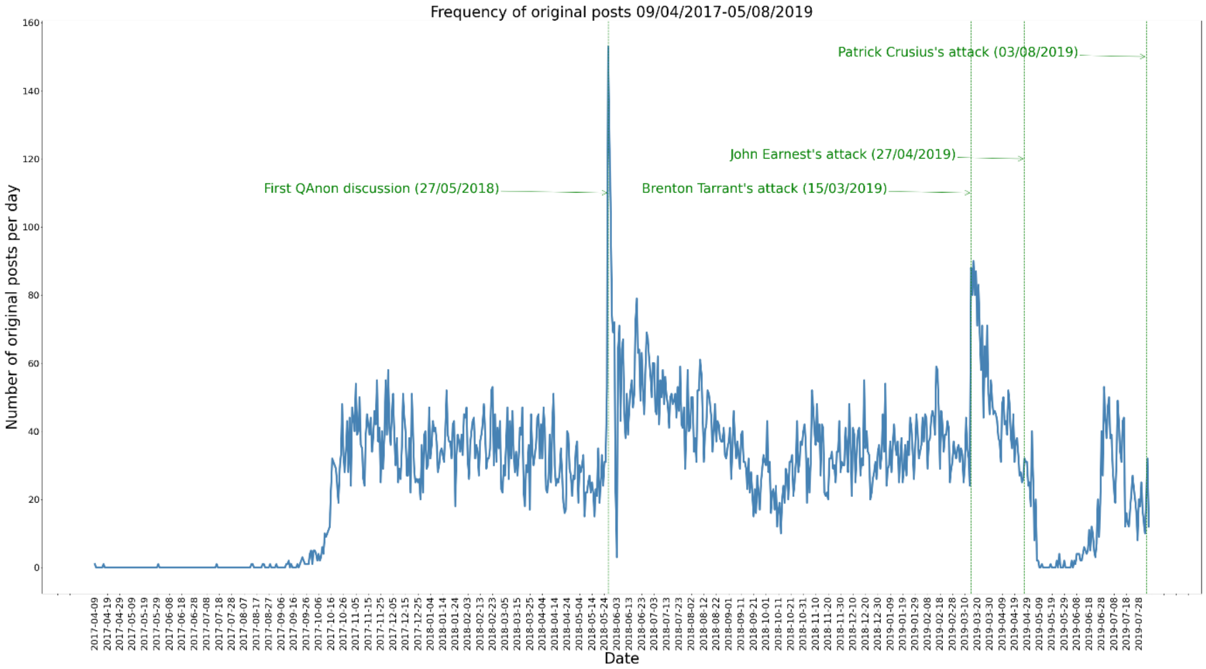

Figure 1: Daily number of original posts made on 8ch.net/pol between 09 April 2017 and 05 August 2019. The green vertical lines indicate dates on which the first QAnon discussion took place and when the RWE attacks that were announced on 8chan/pol occurred, along with the names of the perpetrators. The interested reader can find a full discussion of the behaviour seen here in Baele et al. (2020a).

While Figure 1 provides a measure of the forum's activity, it does not provide an understanding of its content. When working with a large textual data set, such as the one used here, it is useful to develop an overview of the online subculture's ideology, norms, and practices. Fortunately, Python's focus on importable modules makes the implementation of cutting-edge computational analysis techniques, which can help develop such an understanding, easily accessible to researchers.

One such analysis method are the Word2Vec models, as originally formulated by Mikolov et al. (2013). These models are trained on a whole text corpus and produce an *N*-dimensional vector space, whereby each unique word from the text corpus is assigned a distinct vector within this space. The relative positions of these vectors are a product of the contexts in which their corresponding words are used. To put it another way, words that are used in similar contexts are closer together within the vector space than those words that share different contexts of usage. As an example, in a hypothetical corpus and resulting vector space, the vectors for the words "car" and "truck" would be closer together than the vectors for "car" and "apple."

Word2Vec models use artificial neural networks (ANNs), which have gained public attention in recent years due to their use in various projects such as Google Brain and in everyday appliances, such as Amazon's Alexa devices. Essentially, ANNs are intended to mimic the biological process of neurons signalling to one another within a brain.

As Figure 2 depicts, ANNs are made up of nodes, intended to simulate neurons, organised into different layers; the input layer, the hidden layer(s), and output layer. The number of nodes in each layer is largely dictated by the application the ANN is being used for. Each node is connected to other nodes in the "downstream layer," i.e. nodes in the input layer are connected to nodes in the hidden layer. Each node also has a specified threshold value, which is set by the researcher and again has a value that depends on the nature of the project. If the output of a specific node is above its threshold value, it is then activated and sends data to the next layer of the network. If it is not activated, it does not pass on any data. Nodes also have weights assigned to them, which determine how important their signals are, with those that have larger weights contributing more significantly to the output than others. The Word2Vec model uses a shallow network, which are ANNs that have only one or two hidden layers of neurons, like that depicted in Figure 2. In contrast to deep networks, which have many hidden layers. Indeed, this is where the term "deep learning" comes from, with deep learning tasks being those that use deep ANNs.

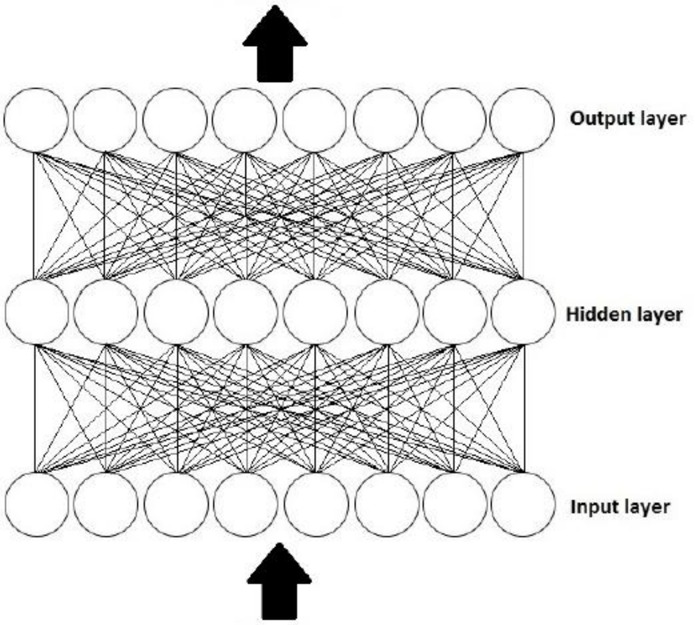

Figure 2: An example of a basic three-layer neural network, where each layer consists of 8 nodes.

Before an ANN is used for whatever application a research project intends, it must first be trained using training data. This involves feeding data into the network, the data being fed through each layer in the network, and the output from the output layer being compared to the initial input data.

The network then uses some form of learning algorithm, the most popular being back propagation (see Rumelhart et al. 1986), to "tune" the nodes in the input weights until the difference between the ANN's output and the initial input data is minimal.

Through such training, ANNs are able to derive patterns and observations from complex data that humans may not be able to, either due to the information within the data appearing to be unrelated or due to the large scale of the data. In the case of textual analysis of online subcultures, the issue is the latter, due to the vast amount of textual data gained from scraping all the posts on sites such as 8chan/pol.

In short, Word2Vec models involve training a shallow neural network on the text corpus, which acts as input training data. However, instead of using the trained network for a specific task, we instead extract the weights from the network's single hidden layer. These extracted weights then constitute the vectors in our N-dimensional space. In other words, these extracted weights are used to determine which words are used in similar contexts based on the text data that the network was trained on.

During training, the MineChan's Word2Vec model[4] was seeded with the most commonly occurring words as dictated by a frequency analysis. The team then asked the model to output the 30 words that were closest in the vector space to each of these seed words. As discussed in Baele et al. (2020a), certain words such as "people," "time," "year," and "thing" were not fed into the model. Such words were not included for one of two reasons. First, because they strongly co-occurred with one of the other, more analytically interesting, terms that were fed in. For example, the words "president" and "trump," which in this case, resulted in the word "trump" being included and "president" excluded. Second, a word was not included if it was a term such as "year" because such terms appeared frequently in the corpus but offer little in terms of analytical value.

While Word2Vec is a powerful analytical tool for working with text data, its multi-dimensional nature provides obvious difficulties in using it to visualise data on a 2D plane, which was its intended usage in Baele et al. (2020a). To ameliorate this, the team utilised the dimensionality reduction technique known as t-Distributed Stochastic Neighbour Embedding (t-SNE), which was originally proposed by van der Maaten and Hinton (2008). In short, this dimensionality reduction method converts high-dimensional data, such as the weights from the hidden layer of our network, into map points for a 2D scatterplot. More specifically, this dimensionality reduction method converts high-dimensional data, $X = \{x_1, x_2 \dots x_n\}$, such as the weights from the hidden layer of our network, into $Y = \{y_1, y_2 \dots y_n\}$, where y_i are map points for a 2D scatterplot. Unlike more traditional, linear, dimensionality reduction techniques that social researchers might be familiar with, such as Principle Components Analysis (PCA), t-SNE utilises a non-linear approach that aims to preserve as much of the high-dimensional data as possible in the resulting low-dimensional map. To achieve this, the algorithm first calculates the corresponding probability of the similarity of the data points in both the high-dimensional and low-dimensional space. This is done by calculating the probability that two points would be neighbours if neighbours were selected in relation to the proportion to their probability density under a Gaussian distribution. The algorithm then tries to minimise the difference between these similarities in higher and lower-dimensional space so as to create an accurate representation in lower-dimensional space.

The output of this whole procedure can be seen in Figure 3, which displays the results of a Word2Vec model implementation of the whole 8chan/pol forum. The output shows thematic clusters that clearly reveal six major components of 8chan/pol's ideology, locating it within the broader "far right" discourse. For example, an orange cluster on the right depicts the predominance of anti-Semitism ("kike," "judaism," "shabbat," "jew," "rothschild," "zionist," etc.).

This cluster is important as "jew" was the third most frequently occurring term of the whole corpus and "jewish" the eleventh. We can also see that the blue cluster consists of terms pertaining to race, with the central grey cluster revealing the importance of discussions revolving around race-oriented movements and action ("identitarian," "organization," "militant," "antifa," "violent," "atomwaffen," etc.), and so on.

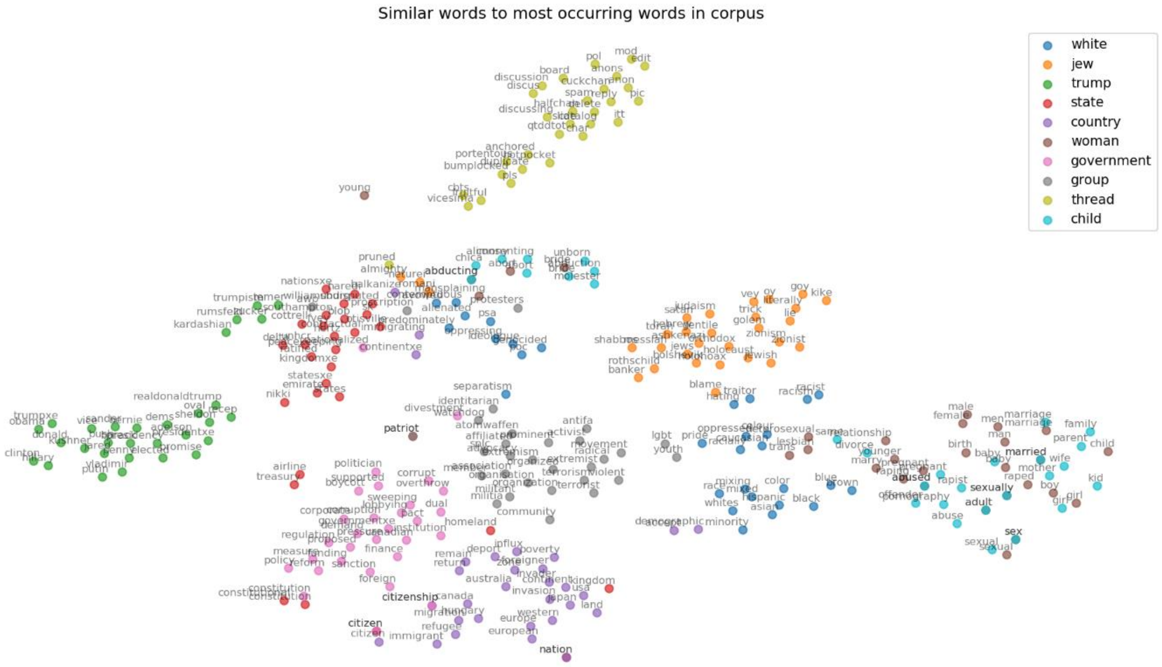

Figure 3: Visualisation of word vectors from the trained Word2Vec model. The legend shows the 10 most frequently occurring words in the 8chan/pol text corpus. The nodes on the graph that match the colour of the words in the legend are then the 30 words that most are most frequently used in the same context as that word.

When studying an online subculture, it is incredibly useful to develop such a "birds-eye view" of the subculture's discussions and ideology as this then enables a more systematic study of specific aspects of the community's online behaviour within the wider context of its overarching ideology. As we have seen, this often entails analysing a large text corpus, which is not realistically possible to do in a systematic manner with traditional social science research methods. However, methods developed in the field of computer science, such Word2Vec models, can be incredibly useful in summarising and visualising a large text corpus in a manner that is easy for a reader to understand. As such, these methods have a clear benefit to investigative social researchers who need to both understand and summarise the themes of a large text corpus.

While the above process may sound quite complex and time consuming to implement, cutting edge computer-assisted analysis techniques, such as Word2Vec models, are easily implemented using a high-level programming language with community developed importable modules. Indeed, Word2Vec models themselves can be easily implemented in Python using either its GenSim or Tensorflow modules, and this is one of the main points that this paper hopes to communicate. The reason why such complex models are easy to implement in Python is twofold. First, packages such as GenSim can be imported into your coding script and contain pre-built implementations of the models that are ready to have data fed into them, in order to train them to produce sensible outputs. All the researcher needs to do is some basic text pre-processing and alter some initial parameters. Second, there is an active online community userbase which enables a researcher to find incredibly helpful tutorials that can take them through the implementation process step-by-step, and which often includes working code examples. In this way, a researcher can utilise such methods without having to have a prior in-depth knowledge of ANNs or coding.

Summary

As society becomes increasingly digital, social scientists are going to have to utilise new data collection and analysis methodologies that allow for large-scale and systematic analysis of online communities and subcultures. Textual data analysis of the discussions that make up such online subcultures allow for this. As we have seen, computer-assisted methods, such as Word2Vec models, enable researchers to develop an overarching view of a community's discussions, which allow for a macro-level understanding of its ideology, norms, and values. The MineChans team often coupled this macro-level understanding with an in-depth qualitative analysis of a stratified sample of threads so as to develop an understanding of the various claims, concepts, ideas, and narratives seen across discussions within the wider context of the over-arching ideology of the subculture.

This combination of methods has not only allowed the researchers to explore the chans online subculture, but to also develop an understanding of the broader far-right online ecosystem; a term that refers to the far-right's presence on all digital platforms, including traditional websites, online forums, and dedicated communication platforms such as Telegram (see Baele et al. 2020; Brace 2021). This has resulted in the work done for the MineChans project being expanded into the ExID project, an international and multi-institutional project that utilises artificial intelligence methods in order to study the broader far-right online ecosystem. This form of analysis has also allowed members of the research team to provide unique insights into the nature of various extremist online communities to numerous law enforcement, security, and legal practitioners. These included highlighting how this online subculture provided clear in-group/out-group and crisis-solution narratives, both of which are present in almost all extremist ideologies (Berger 2018a, 2018b).

The key point that this paper has sought to make, however, is that these computational methods offer social scientists a lot in terms of new analytical capabilities, particularly in the digital age, and that they are therefore worth exploring. This is especially true for researchers looking at online communities. Furthermore, the nature of high-level programming languages, such as Python and its accompanying importable modules, mean that the implementation of these methods is now much easier for researchers with relatively limited knowledge of computer coding than they once were.

Funding

This work was supported by Economic and Social Research Council (ESRC) via the Centre for Research and Evidence on Security Threats (CREST) consortium under [Grant number ES/N009614/1].

Useful Resources

Social researchers are likely to benefit from the fact that Python is currently one of the most popular languages used by social data scientists. This means that there is a vast collection of free, open-source, tutorials available to those wanting to learn Python. Two examples of this are the "30 Days of Python" course by the GitHub user Asabeneh and the Python materials available here, maintained by the Q-Step Centre at the University of Exeter, UK. An incredibly useful tutorial on how to implement a Word2Vec model using Python's GenSim package can also be found here.

Notes

[1] For more information on the MineChans project go to this web page.
[2] Unfortunately, the source of this information cannot be disclosed.
[3] Details about the Twitter API can be found here.
[4] There are different "flavours" of Word2Vec models, with the MineChans project using the skip-gram version. Unfortunately, there is not space here to discuss the subtle differences in these family of models. The interested reader is advise to visit this link, which gives a good overview of the different implementations.

Bibliography

Baele, S., Brace, L., and Coan, T. 2020a. "The 'Tarrant Effect': What Impact did Far-Right Attacks Have on the 8chan Forum?" Behavioral Sciences of Terrorism and Political Aggression. doi:10.1080/19434472.2020.1862274.

Baele, S., Brace, L., and Coan, T. 2020b. "Uncovering the Far-Right Online Ecosystem: An Analytical Framework and Research Agenda." Studies in Conflict & Terrorism. doi: 10.1080/1057610x.2020.1862895.

Baele, S., Brace, L., and Coan, T. 2021. "Variations on a Theme? Comparing 4chan, 8kun, and Other chans' Far-Right "/pol" Boards." Perspectives on Terrorism, 15 (1): 65-80. https://www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2021/issue-1/baele-et-al.pdf.

Berger, J. 2018a. Extremism. Cambridge, MA, USA: MIT Press.

Berger, J. 2018b. "The Difference Between a Killer and a Terrorist." The Atlantic, April 26. https://www.theatlantic.com/politics/archive/2018/04/the-difference-between-killer-and-terrorist/558998/.

Blevins, K. and Holt, T. 2009. "Examining the Virtual Subculture of Johns." Journal of Contemporary Ethnography 38 (5): 619-648.

Brace, L. 2021. "The Role of the Chans in the Far-Right Online Ecosystem." Global Network on Extremism & Technology, April 1. Accessed April 2, 2021. https://gnet-research.org/2021/04/01/the-role-of-the-chans-in-the-far-right-online-ecosystem/.

Conway, M. 2016. "Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research." Studies in Conflict & Terrorism 40 (1): 77-98.

Conway, M., Scrivens, R., and Macnair, L. 2019. "Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends." ICCT Policy Briefs, November 25. https://icct.nl/publication/right-wing-extremists-persistent-online-presence-history-and-contemporary-trends/.

Crawford, B. 2020. "/K/ and the Visual Culture of Weapons Boards." Centre for Research and Evidence on Security Threats, October 28. Accessed June 16, 2021. https://crestresearch.ac.uk/comment/k-and-the-visual-culture-of-weapons-boards/.

Davey, J. and Ebner, J. 2017. "The Fringe Insurgency – Connectivity, Convergence and Mainstreaming of the Extreme Right." Institute for Strategic Dialogue, October. London, UK: ISD.

DiMaggio, P. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." Big Data & Society. doi: 10.1177/2053951715602908.

Edelmann, A., Wolff, T., Montagne, D., and Bail, C. 2020. "Computational Social Science and Sociology." Annual Review of Sociology 46: 61-81. doi: 10.1146/annurev-soc-121919-054621.

Europol. 2020. "European Union Terrorism Situation and Trend Report (Te-Sat) 2020." June 23. Accessed August 1, 2020. https://www.europol.europa.eu/activities-services/main-reports/european-union-terrorism-situation-and-trend-report-te-sat-2020.

Evans, R. 2019a. "Ignore the Poway Synagogue Shooter's Manifesto: Pay Attention to 8chan's /pol/ Board." Bellingcat, April 28. Accessed May 5, 2019 https://www.bellingcat.com/news/americas/2019/04/28/ignore-the-poway-synagogue-shooters-manifesto-pay-attention-to8chans-pol-board/.

Evans, R. 2019b. "Shitposting, Inspirational Terrorism, and the Christchurch Mosque Massacre." Bellingcat, March 15. Accessed April 7, 2019. https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/.

Evans, R. 2019c. "The El Paso Shooting and the Gamification of Terror." Bellingcat, August 4. Accessed September 13, 2019. https://www.bellingcat.com/news/americas/2019/08/04/the-el-paso-shooting-and-the-gamification-of-terror/.

Fernandez, M., Asif, M., and Alani, H. 2018. "Understanding the Roots of Radicalisation on Twitter." WebSci '18: 10th ACM Conference. doi: 10.1145/3201064.3201082.

Haenfler, R. 2013. Subcultures: The Basics. London, UK: Routledge.

Herbert, S. 1998. "Police Subculture Reconsidered." Criminology 36: 343-369.

Holt, T. and Copes, H. 2010. "Transferring Subcultural Knowledge Online: Practices and Beliefs of Persistent Digital Pirates." Deviant Behaviour 31: 625-654.

Horgan, J. 2014. The Psychology of Terrorism. Abingdon, UK: Routledge.

Keen, R., Crawford, B., and Suarez-Tangil, G. 2020. "Memetic Irony and the Promotion of Violence Within Chan Cultures." Centre for Research and Evidence on Security Threats, December 15. Accessed June 16, 2021. https://crestresearch.ac.uk/resources/memetic-irony-and-the-promotion-of-violence-within-chan-cultures/.

Kruglanski, A., Bélanger, J., and Gunaratna, R. 2019. The Three Pillars of Radicalization: Needs, Narratives, and Networks. Oxford, UK: Oxford University Press.

Lazer, D., Pentland, A., Watts, D., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M., Strohmair, M., Vespignani, A., and Wagner, C. 2020. "Computational Social Science: Obstacles and Opportunities." Science 369 (6507):1060-1062.

Ludemann, D. 2018. /pol/emics: Ambiguity, Scales, and Digital Discourse on 4chan. Discourse, Context & Media 24: 92-98.

Massanari, A. 2017. "#Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures." New Media & Society 19 (3): 329-346.

McCauley, C. and Moskalenko, S. 2017. Friction: How Conflict Radicalizes Them and Us. Oxford, UK: Oxford University Press.

Merrin, W. 2019. "President Troll: Trump, 4chan and Memetic Warfare." In Trump's Media War, edited by Happer, C., Hoskins, A., and Merrin, W., 201-226. Cham, Switzerland: Palgrave Macmillan.

Mikolov, T., Corrado, G., Chen, K., and Dean, J. 2013. "Efficient Estimation of Word Representations in Vector Space." Proceedings of the International Conference on Learning Representations (ICLR 2013).

Nagle, A. 2017. Kill All Normies: Online Culture Wars from 4chan and Tumblr to Trump and the Alt-Right. Winchester, UK: Zero Books.

Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R., and Winters, J. 2020. "How We Do Things With Words: Analyzing Text as Social and Cultural Data." Frontiers in Artificial Intelligence. doi: 10.3389/frai.2020.00062.

O'Connor, B., Bamman, D., and Smith, N. 2011. "Computational Text Analysis for Social Science: Model Assumptions and Complexity." Proceedings of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds.

Reeve, Z. 2019. "Terrorist Psychology and Radicalisation." In Routledge Handbook of Terrorism and Counterterrorism, edited by Silke, A., 125-134. Abingdon, UK: Routledge.

Rumelhart, D., Hinton, G., and Williams, R. 1986. "Learning Representations by Back-Propagation Errors." Nature 323 (6088): 533-536.

Trammell, M. 2014. "User Investment and Behavior Policing on 4chan." First Monday. https://journals.uic.edu/ojs/index.php/fm/article/download/4819/3839.

Van der Maaten, L. and Hinton, G. 2008. "Visualising Data Using t-SNE." Journal of Machine Learning 9: 2579-2605.

Zamani, M., Rabbani, F., Horicsányi, A., Zafeiris, A., and Vicsek, T. 2019. "Differences in Structure and Dynamics of Networks Retrieved from Dark and Public Web Forums." Physica A 525: 326-336.

Zhang, J., Wang, W., Xia, F., Lin, Y., and Tong, H. 2020. "Data-Driven Computational Social Science: A Survey.' Big Data Research 21. doi: 10.1016/j.bdr.2020.100145.