Computers in Human Behavior 116 (2021) 106646

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Faces of radicalism: Differentiating between violent and non-violent radicals by their social media profiles

Michael Wolfowicz *, Simon Perry, Badi Hasisi, David Weisburd

Institute of Criminology, Faculty of Law, and the Cyber-Security Research Centre, Hebrew University of Jerusalem, Israel

ARTICLE INFO

Keywords:
Terrorism
Case-control
Social-media
Internet
Social-learning theory

ABSTRACT

Objectives Social media platforms such as Facebook are used by both radicals and the security services that keep them under surveillance. However, only a small percentage of radicals go on to become terrorists and there is a worrying lack of evidence as to what types of online behaviors may differentiate terrorists from non-violent radicals. Most of the research to date uses text-based analysis to identify "radicals" only. In this study we sought to identify new social-media level behavioral metrics upon which it is possible to differentiate terrorists from non-violent radicals. Methods: Drawing on an established theoretical framework, Social Learning Theory, this study used a matched case-control design to compare the Facebook activities and interactions of 48 Palestinian terrorists in the 100 days prior to their attack with a 2:1 control group. Conditional-likelihood logistic regression was used to identify precise estimates, and a series of binomial logistic regression models were used to identify how well the variables classified between the groups. Findings: Variables from each of the social learning domains of differential associations, definitions, differential reinforcement, and imitation were found to be significant predictors of being a terrorist compared to a nonviolent radical. Models including these factors had a relatively high classification rate, and significantly reduced error over base-rate classification. Conclusions Behavioral level metrics derived from social learning theory should be considered as metrics upon which it may be possible to differentiate between terrorists and non-violent radicals based on their social media profiles. These metrics may also serve to support textbased analysis and vice versa.

1. Introduction

The social media activities of radicals ought to provide an important window of opportunity to identify potential terrorists before they attack (Gill et al., 2017; Pelzer, 2018). Yet from New York to Paris and Jerusalem, we are too often reminded of those terrorists who manage to 'fly under the radar' (Quiggin, 2017). The literature reveals the existence of a number of fundamental issues in the research underpinning current approaches to identifying terrorists based on the surveillance of social-media.

The issues already begin at the unit of analysis, and a tendency to conflate "radicals" with terrorists, which are actually distinct groups (Wolfowicz, Litmanovitz, Weisburd, & Hasisi, 2020). Radicals are individuals who support or justify terrorism, and in some cases express a willingness to engage in acts of radical violence. However, less than 1% of radicals will ever move from this state of cognitive radicalization to the actual carrying out of terrorism offences, with the remainder remaining forever inert (McCauley & Moskalenko, 2017). Most studies in the area of automated detection of "radicalization" are actually presenting methods for the identification of radicals from among the general population; not terrorists. Some studies even make the false assumption that their findings ought to be transferable to terrorists, who are wrongfully conceived as simply being highly radicalized radicals (Pelzer, 2018).

This assumption gives rise to a serious statistical issue as it pertains to the use of common, text-based analysis for the automated detection of "radicalization". If radicals and terrorists are using the same lexicon, and radicals represent a much larger group than terrorists, then the majority of instances of radical language that such methods identify belong to non-violent radicals (Parekh, Amarasingam, Dawson, & Ruths, 2018; Shortland, 2016). Given that the pool of radicals who will remain forever inert is significantly larger than the pool of radicals who will eventually become terrorists, text-based approaches will result in an unacceptable false-positive rate (Neuman, Cohen, & Neuman, 2019).

Terrorism is a rare event and identifying potential terrorists before they strike is like finding the proverbial needle in the haystack (Neuman

* Corresponding author. Institute of Criminology, Faculty of Law, Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel.
E-mail address: michael.wolfowicz@mail.huji.ac.il (M. Wolfowicz).

https://doi.org/10.1016/j.chb.2020.106646
Received 27 June 2020; Received in revised form 22 November 2020; Accepted 28 November 2020
Available online 7 December 2020

et al., 2019). As has been pointed out in the literature, while it is certainly informative to identify how radicals and terrorists differ from the general population, we are ultimately most interested in identifying terrorists from among the larger radical population (Freilich & LaFree, 2016). As such, when it comes to leveraging social media as a site for surveillance and the identification of terrorists, it is first necessary to identify what types of online behaviors differentiate the small number of terrorists from the large pool of non-violent radicals. The lack of this type of inquiry represents a significant gap in the body of knowledge (Neuman et al., 2019; Schmid & Forest, 2018; Scrivens, Gill and Conway, 2019).

A third problem that exists pertains to the focus and objectives of much of the current research. To a large degree the literature has been overly focused on the performance of different algorithms, and studies have failed to refer, *a priori,* to how they are informed by, or inform theoretical models. This has led to issues concerning what factors should be being examined in the first place (Pelzer, 2018; Settanni, Azucar, & Marengo, 2018). As noted above, most studies employ some form of text-based analysis, which seek to identify the usage of key words or phrases associated with radical lexicon. However, trends in social-media use have changed considerably in recent years, and text-based posts represent only a fraction of users' social media output. As such, text-only based analyses may reach false conclusions about the individual being analyzed.

In order to address these issues, it is imperative that we identify theory-driven, social-media level metrics upon which we may better differentiate terrorists from the wider pool of non-violent radicals. In this study, we draw on an established criminological framework, social learning theory, to test whether behavioral level metrics can effectively discriminate between terrorists and non-violent radicals based on their Facebook profiles. Using a matched case-control design, we compare the Facebook activities and interactions of 48 Palestinian terrorists in the 100 days leading up to their attack with a matched sample of non-violent radicals.

1.1. Possible metrics for differentiating terrorists from non-violent radicals

To date, the most popular approach to the automated detection of 'radicalization' or 'extremism' has been text-based analysis, which seeks to identify patterns in the use of pre-determined keywords and phrases that are known to be commonly associated with radicalization (Pelzer, 2018). This may include words such as "martyr" or "jihad", or phrases such as "heil hitler" and different variations of their use. But similar terms could just as easily be used by opponents who are disparaging radicals or radical ideology. As such, to improve the accuracy of text-based analysis, studies have integrated advanced forms of sentiment and semantic analysis (Scrivens, Davies, & Frank, 2018; Scrivens, Gaudette, Davies and Frank, 2019). Sentiment analysis identifies whether the emotions associated with the usage of the words or phrases are more negative, positive, or neutral, thereby discriminating between positive and negative uses of the key terms (Ortigosa, Martin and Carro, 2014). Semantic analysis seeks to identify and incorporate the meaning of a word, and the distance between different forms of its usage, improving the accuracy of detecting sentiment (Bogolyubova, Panicheva, Tikhonov, Ivanov, & Ledovaya, 2018).

While some believe that approaches combining sentiment and semantic analysis is "the future" of detecting radicalization online, these approaches suffer from a number of limitations (Scrivens et al., 2019). Firstly, sentiment analysis is quite limited in the range of emotions they can capture, as well as their capacity to account for context and features that are difficult to identify, such as sarcasm. Secondly, depending on the issue being examined, positive and negative sentiments do not always translate as good and bad respectively (Gaspar, Pedro, Panagiotopoulos, & Seibt, 2016). These issues may serve to explain why security services' use of such approaches to identify terrorists have been shown to have high false-positive rates, leading to many false arrests (Hasisi, Perry, & Wolfowicz, 2019). False-arrests can potentially contribute to stigmatization and perceived injustice. As a result, this can contribute to increasing the likelihood of a backlash effect that includes terrorism (Tankebe, 2020). To help mitigate such limitations, even highly advanced sentiment-based approaches should include human verification (Chatterjee et al., 2019; Gaspar et al., 2016).

Another way to improve these approaches may be to incorporate 'weak signals', or proximal indicators of increasing risk that have been identified using traditional risk-assessment approaches. Two of these signals, identification and fixation, have demonstrable usefulness. Identification can be observed in a user's online behaviors in which they express aspirations to be like the "pseudo-commando" represented by terrorists from a range of ideologies. Indicators of increasing identification may be seen in online postings that demonstrate an affinity for weapons and terrorists who have already carried out attack. Fixation relates to an increasing 'pathological preoccupation' with a radical ideology or cause. Like normative individuals, radicals also make online posts about innocuous personal experiences and events. Posts that pertain to their radical ideology may represent only a small proportion of all posts. As such, when the proportion of an individual's posts pertaining to their radical ideology increase, this is a sign of increasing fixation (Brynielsson et al., 2013; Cohen, Johansson, Kaati, & Mork, 2014; Kaati, Shrestha, & Cohen, 2016).

For all of their potential, the approaches described above are still limited to the analysis of user-produced text. This is a significant limitation given that user-generated text-based posts represent only one, specific source of information among the plethora of information pertaining to an internet user. For example, trends in social-media usage have seen a decrease in text-based posts, which have increasingly been replaced with the posting of images, videos, and shared content. Even on Twitter, a traditionally text-based platform, users are writing less and sharing more graphical content. On platforms like Facebook, images now represent the bulk of communicative user-generated posts. Given that images posted on Facebook may be highly predictive of personality (Eftakhar, Fullwood and Morris, 2014), text-only based approaches are missing important information. While some advanced applications of sentiment analysis for multimedia have been developed, they are still in their infancy and have yet to be fully integrated in multimodal applications (Li, Fan, Jiang, Lei, & Liu, 2019). Qualitative analysis carried out by humans may still be superior for classifying the sentiment in multimedia (Gaspar et al., 2016).

The literature suggests that one way that automated detection can be improved is by incorporating behavioral factors, sometimes referred to as "digital footprints" or 'non-explicit' factors. Behavioral factors include "activities", such as posting frequency and type, and "interactions" such as the receiving of likes (e.g. Pressman & Ivan, 2016). Additional metrics of this variety that have recently been examined include: post type (owner-created vs. owner-shared), network size (number of friends), and the ratio between positive and negative posts (Ophir, Asterhan and Shwarz, 2019). These types of factors have been shown to be predictive of a range of cognitions and offline behaviors (Chancellor & De Choudhury, 2020; Ophir, Asterhan and Shwarz, 2019).

Unfortunately, only a small number of studies have incorporated such factors as predictors of radicalization on social media. Ferrara, Wang, Varol, Flammini, and Galstyan (2016) found that the ratio between originally authored and shared posts played the most important role of all factors in predicting the adoption of radical lexicon. Sutch and Carter (2019) found that post frequency and account duration predicted the use of extremist lexicon. Posting frequency and network size have also been found to increase the predictive quality of text-based identification of online extremism (Nouh, Nurse, & Goldsmith, 2019). In what is arguably the most comprehensive application of such factors to date, Smith, Blackwood, and Thomas (2020) analyzed differences in profile-level behaviors between ordinary Twitter users and users

identified as being supporters of ISIS (Islamic State in Iraq and Syria). The study examined metrics such as the ratio between originally authored and re-tweeted posts made by the users, how long the account had been open, posting frequency, and network size. All of these factors were found to be predictive of membership in the ISIS supporting group.

Each of these behavioral level factors have been shown to be able to differentiate between radicals and the general population. That is, they have been shown to be predictive of individuals who are presumably cognitive radicals who support or justify terrorism. What we still don't know is how these factors may serve to differentiate between non-violent radicals and terrorists. Nevertheless, there is good reason to believe they can, as each of these metrics can be seen to be analogous to dimensions of one of the most empirically proven criminological frameworks, social learning theory.

1.2. Operationalizing online behaviors: A social learning framework

According to Social Learning Theory, the learning of deviant behaviors occurs through the same dynamics as the learning of normative behaviors, namely 1) differential associations, 2) definitions, 3) imitation, and 4) differential reinforcement. Differential associations are figures from an individual's network, such as peers, parents, other models, and media. Differential associations provide the individual with a balance of definitions in support of, or against a given behavior. When the balance of definitions in favor of the behavior outweigh those against it, the individual is at a heightened likelihood of engaging in that behavior. Differential associations also serve as a source of imitation, and provide differential reinforcement with respect to a given behavior (Akers & Sellers, 2004; Akers & Silverman, 2014; Akins & Winfree, 2016).

There is quite a bit of evidence to support the relevance of the social learning framework to understanding the social media-radicalization nexus. Increased frequency of social media usage increases the likelihood that a user will come into contact with radical content (Costello, Hawdon, Ratliff, & Grantham, 2016). Exposure to radical content online, especially when it is actively sought after, increases the likelihood that an individual will hold radical attitudes (Frissen, 2020). Additionally, being online friends with other radicals online has been found to increase the likelihood that an individual will hold radical attitudes (Wojcieszak, 2008, 2010; Pauwels & Schils, 2016). Moreover, both active and passive forms of radical content consumption have been found to be predictive of sub-terrorists forms of radical behaviors, such as ideologically motivated attacks on persons and property (Pauwels & Schils, 2016; Pauwels & Hardyns, 2018).

Beyond this, social learning perspectives also posit reciprocal determinism (Bandura, 1986). Reciprocal determinism is highly relevant to the internet-radicalization nexus as it suggests that an internet user's cognitions and behaviors influence their internet usage, and their online experience influences their cognitions and behaviors in a recurring cycle (Frissen, 2020). Evidence of reciprocal determinism was found by Ness et al. (2017) who found that pre-existing ideological attitudes were reinforced when exposed to ideological congruent websites. More importantly, in line with reciprocal determinism, internet behaviors on platforms such as Facebook can be both an influencer and predictor of a range of offline deviant behaviors (D'Angelo, Kerr and Moreno, 2014; Frost & Rickwood, 2017; Kingston, Fedoroff, Firestone, Curry, & Bradford, 2008).

As such, while translating social learning variables as online behaviors and interactions represents a divergence from the original theory, it is also intuitive. The literature has consistently found that online behaviors and interactions derived from social learning theory are predictive of offline offending behaviors. For example, having online deviant peers who post about their own offline criminal behaviors, increases the likelihood of the receiving user engaging in the same behaviors (Pratt et al., 2010; McCuddy & Vogel, 2015a, 2015b), including sub-terroristic radical violence (Pauwels & Schils, 2016). Like offline, online differential associations are conditioned by frequency and duration, which can also relate to the engagement with a particular media (Bandura, 1978), and network size (Haynie, 2001, 2002; Haynie, Doogan, & Soller, 2014). Such factors have been found to be statistically significant predictors for a range of offline criminal behavior (McCuddy & Vogel, 2015a, 2015b).

Definitions, or the attitudes that an individual hold toward a given behavior (Akers, 1998), are generally assessed through survey instruments. In the context of radicalism and terrorism, definitions can be assessed by the extent to which an individual identifies with an ideology, group or cause (Smith, Blackwood, & Thomas, 2020). Definitions may therefore be observed in online posting behaviors as indicating a degree of fixation when a greater proportion of the user's content output focuses on the radical topic. Fixation may be a 'weak signal' for the move from radical attitudes to behavior (Brynielsson et al., 2013; Reid Meloy, Hoffmann, Guldimann, & James, 2012).

The additional elements of social learning theory, imitation and differential reinforcement are rarely examined, partly due to difficulties in measurement. However, the internet offers unique ways in which they may be measured (Holt & Bossler, 2016). The variety of approaches taken (e.g. Holt, Burruss, & Bossler, 2010; Shadmanfaat et al., 2020; Skinner & Fream, 1997) highlights that imitation refers not only to the modelling of a specific criminal behavior, but the behaviors of differential associations more generally (Akers, 1998). On social media platforms, the sharing of content is said to represent a form of such imitation (Hong & Gardner, 2014). Sharing may indicate a user's emotional connectedness with the content, which may better express a shared opinion than what the individual is capable of in their own words (Kim & Yang, 2017).

Differential reinforcement relates to the anticipated approval that one will receive from their peers and network for engaging in a specific behavior. Receiving of differential reinforcement, or observing that which others receive can encourage engagement in the behavior. Online peer approval has been found to correlate with a range of cyber-deviant behaviors (Holt et al., 2010; Miller & Morris, 2016). Measured by looking at the reactions garnered by users' posts (e.g. likes, comments, shares etc.), differential reinforcement has been found to be predictive of dangerous cognitive states (Brown et al., 2019; Hussain et al., 2019). Yet, differential reinforcement is rarely examined, including with respect terrorism (Cone, 2016; Shapiro & Maras, 2019).

The operationalization of behavioral metrics drawn from traditional criminological theory is an innovative approach to understanding which factors may be important predictors for offline behaviors and why (Pelzer, 2018; Scrivens, Davies, & Frank, 2020). Based on social learning's reciprocal determinism, we can understand how and why online behaviors analogous to components of social learning have been found to be significant factors for differentiating radicals from the general population, and why they ought to be predictive of offline behaviors. Examining how such metrics may differentiate perpetrators of offline violence, such as terrorism, from a pool of potential offenders, such as non-violent radicals, addresses an important gap in the literature (Patton et al., 2014). In seeking to examine a range of social learning derived metrics (Holt et al., 2010), we accept that the variables operate cumulatively and interactively (Peterson & Densley, 2017). It is in the context of this framework that we seek to identify whether behavioral and interaction level metrics on Facebook differentiate between terrorists and non-violent radicals. In accordance with the theoretical framework, we hypothesize that higher values on variables from each of the social learning domains increases the likelihood of being a terrorist over a non-violent radical.

2. Methods

2.1. Context

To date, only a small number of studies have compared the online

behaviors of terrorists—usually foreign fighters—with non-violent radicals. However, these analyses have been based on exceptionally small datasets, and limited to examining factors identifiable in lexicon (Dillon, Neo and Frelich, 2019, pp. 1–24; Seng, Khader, & Pang, 2018, p. 87). The lack of evidence is understandable given the low base-rates of terrorism offending (Gill, Horgan, Corner, & Silver, 2016). As such, researchers have often turned to the Israeli context as a case study, where relatively elevated base rates of terrorism offending provide for the ability to carry out meaningful statistical analysis (Hasisi, Carmel, Weisburd, & Wolfowicz, 2019).

Beyond providing for an effective 'laboratory', the Israeli case is also demonstrative of why it is important to identify metrics that can effectively discriminate between terrorist and non-violent radicals online. Like other countries, Facebook remains the most widely used platform among Palestinians. During the terrorism wave in Israel from 2014 to 2018, officials often implicated Facebook as one source of the spread of radicalization. They also reported that for a number of terrorists, clear signs of radicalization and even intent were identifiable in their Facebook profiles. Israeli authorities subsequently carried out a number of arrests based in part on the analysis of Palestinians' Facebook behaviors. While Israeli authorities claimed that this approach successfully prevented numerous, additional attacks, reports also indicated that this approach led to hundreds of false-positives, many of which resulted in false arrests (Hasisi, Perry, & Wolfowicz, 2019).

As discussed above, one of the challenges to differentiating between non-violent radicals and terrorists based on what they post online is that the non-violent radical population is so much larger than the terrorist population. Due to the differences in population size and a shared lexicon, the majority of radical postings are being made by non-violent radicals, further reducing the likelihood of successfully identifying a terrorist (Shortland, 2016). This Israeli case therefore poses an even greater challenge. According to the 2008 European Values Survey, 5.3% of participants responded that terrorism can sometimes be justified (EVS, 2008). This figure is only slightly larger than the 4.6% of Western European respondents to the 2007 Pew Report who stated that suicide bombings were often justifiable. Although an additional 10% responded that they could occasionally be justified, in contrast, 47% of Palestinian respondents in the same survey stated that suicide bombings were often justifiable, and an additional 33% that it was occasionally justified (PEW, 2007). According to a survey carried out by the Jerusalem Media and Communications Center—a Palestinian entity—in December 2014, 78% of respondent expressed their support for the "increase in Jerusalem and the rest of the West Bank in attempts to stab or run over Israelis.", referencing what was then the start of the years long terrorism wave (JMCC, 2014).

Relative to other platforms, digital trails left on Facebook are considered to be accurate reflections of offline personalities, emotions and psychological states and can be used to predict a range of offline deviant behaviors (e.g. Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011; Marder, Joinson, Shankar and Houghton, 2016). Theoretically, digital trails from Facebook should be able to predict differential radical outcomes (Bartlett and Reynolds, 2015). As such, in this study we sought to address two research questions. First, can metrics derived from a social learning framework, as applied to Facebook user-level behavior and interactions, serve to differentiate between terrorists and non-violent radicals? Secondly, what level of classification can be achieved by a model based solely on these behavior and interaction level factors? We hypothesize that higher scores on each of the social learning metrics increases the likelihood of being a terrorist over a non-violent radical. We also hypothesize that behavior and interaction level metrics will provide for an improvement over the base-rate level of classification.

2.2. Data

Between 2016 and 2018 a team of Arabic speaking researchers built upon an existing database of lone actor terrorists who had carried out ideologically/politically motivated attacks in Israel between 2014-2018.[1] The research team scoured Facebook for the open and public profiles of individuals in the database using software provided by the Israeli intelligence company Terrogence. During data collection, as new attacks occurred, the team searched for the profiles of the attackers and added to the database when possible. From a total sample of 150 terrorists, it was known that least 100 of them had Facebook at some point, based on screenshots published in both Israeli and local Arabic media. In total, our extensive searches identified 60 of the terrorists' Facebook profiles, of which 48 were fully public and open access profiles.[2] As such, our sample includes a large proportion of the known Facebook profiles for terrorists who were known to have carried out attacks during the observation period, and we believe that this sample provides for a high degree of representativeness. The 48 terrorists included in the sample had been responsible for carrying out a range of terrorist attacks, including stabbing/bladed weapon attacks (N = 24), driving/vehicular attacks (N = 9), firearm attacks (N = 4), combined attacks (N = 4) attacks using explosives (N = 3), and other (N = 4) attacks. All of these attacks were listed as incidents of terrorism, defined as ideologically/politically motivated violence, by the Israel Security Agency (ISA), the country's domestic intelligence and security service.

2.2.1. Control group selection: matching procedure

In the current study we sought to identify a control or comparison group of non-violent radicals for our identified sample of terrorists. Given the limitations of retrieving Facebook data,[3] such as the absence in many cases of key matching variables such as age and location,[4] it was not possible to use a traditional matching approach (e.g Propensity Score Matching). As such, to construct a valid control group, we devised a matching strategy that mimics the principles of statistical matching by leveraging Facebook's Graph Search (Minkus, Ding Dey and Ross, 2015).

Facebook's Graph Search allows the development of a valid comparison group by identifying similar types of users. Results are prioritized by those closest to each other according to the selected criteria (e.g., age, sex, location) (Curtiss et al., 2013). Compared to other methods, Facebook's Graph Search is biased in prioritizing higher-degree nodes (Gjoka, Kurant, Butts, & Markopoulou, 2010; Kurant, Markopoulou, & Thiran, 2010), profiles that have interacted with each other more, through likes, comments, etc. The search "computes a floating point score" for each search query, taking into account these and other factors, and multiplies the default results by a constant factor, and then: "The index server returns the results with the highest scores", with aggregators giving "priority to documents with higher scores" (Curtiss et al., 2013, p. 5).

[1] As part of the EU funded PRIME project. All attacks had been classified as acts of terrorism by the Israel Security Agency (ISA). See: http://www.fp7prime.eu/.

[2] It should be considered that terrorist profiles are systematically removed by Facebook following attacks (sometimes by the request of authorities), or by family members' intervention.

[3] Data scraping of Facebook is against Facebook's terms and conditions.

[4] Even among open profiles, characteristics such as age and location may be made private. For example, van Dam and Van De Velden (2015) found that among 43,861 Facebook users, while gender was available for over 99% of the sample, location was only available for 54.2%, hometown for 29.8%, friends list for 48.6%, and date of birth was only available for 2.5%. As Farahbakhsh, Han, Cuevas, and Crespi (2013) found, even among public profiles, key characteristics of the users are not always made available. Results from search queries that include such factors are limited to profiles in which such information is available to Facebook, even if it is not publicly visible.

To maximize the use of the terrorists' data, we sought a 1:2 match for each terrorist case. We employed a search string of "MALES who are friends of *USERID*, aged *AGE*, from *CITY*[5] on each of the cases' profiles. Results were then reviewed manually and maintained only when they contained at least one radical post[6] (Smith, Wakeford, Cribbin, Barnett, & Hou, 2020). For 23 cases only two matches were found that met these criteria. For cases with more than 2 eligible matches (N = 12), selection was made based on additional criteria such as; attending the same school (N = 4), sharing the same tribal name (N = 5), or appearing in a publicly available picture with the case (N = 3). In one case there were no additional features upon which to select, so profiles were chosen at random. The remaining 12 cases had hidden friend lists, so we compiled a list of users who had liked or commented on their posts and manually applied the selection criteria, selecting the first two successful matches. By including relationships and place of residence as matching variables, the matches inherently account for multiple unmeasured and unmeasurable factors, resulting in "genuinely matched data" (Kuo, Duan, & Grady, 2018, p. 10). The final dataset consisted of N = 48 terrorists and N = 96 non-violent male radicals, nearly identical in age, and originating from just 21 cities, towns, or villages (Table 1) (Table 2).

Based on the assumption that the overwhelming majority of radicals will remain inert, and less than 1% of radicals will ever go on to become terrorists (McCauley & Moskalenko, 2017), our control group should be considered a sample of true non-violent, cognitive radicals, and not as potential terrorists.

2.2.2. Time frame

Given average timeframes for the move from radical beliefs to behaviors (e.g. Klausen, Libretti, Hung, & Jayasumana, 2018), and timeframes used in prior research (e.g. Ophir, Asterhan, & Schwarz, 2019), all profiles were coded manually from the day of the attack (day "0″) up to day 100 before the attack. The control group profiles were therefore coded for the same date range as their matched cases. This additional level of matching helped to deal with contamination in the form of the terrorists' actions influencing the posting behaviors of the controls. Importantly, it also helped to control for temporally-dependent environmental conditions, such as the general socio-political atmosphere at the time, and specific events. Since the matched pairs also live in the same locale, this also helps to control for more localized ecological factors, such as the occurrence of similar events (Phillips, Matusko, & Tomasovic, 2007).

2.3. Variables

The dependent variable was a dichotomous variable in which all cases in the terrorist sample were coded as "1″, and all cases from the control group of non-violent radicals were coded as "0". Based on prior research, we translate behavioral level metrics as dimensions of the social learning framework. We coded *differential associations* as a dichotomous variable in which 1 = when a case posted about a prior terrorist attack committed by a Facebook friend. For *definitions*, the research team used a quantitative content analysis (QCA) tool they developed based on Holbrook & Taylor (2014) to assess whether a post was "radical". While the tool included a 0–3 scale, for this study the variable was dichotomized to represent either radical (1) or non-radical (0) content. We subsequently constructed a measure for the proportion of all posts that were 'radical'.[7] *Imitation* was based on post type, in which we calculated the number of posts that were; image posts, video posts, text posts, and shared posts. Following prior studies (e.g. Smith, Wakeford, et al., 2020), a ratio was created between originally authored text-based posts and shared posts. For *differential reinforcement*, we calculated the average number of likes, comments, and shares received per post[8,9].

Table 1
Characteristics of the sample.

Variable	Terrorists (N = 48)	Non-violent (N = 96)
Age	21.125 (SD = 4.579)	21.125 (SD = 4.554)
Male	**100%**	**100%**
Jerusalem	25%	25%
West Bank	75%	75%
Cities	30.6%	30.6%
Villages	69.4%	69.4%
At least 1 radical post	98%	100%

Table 2
Descriptive statistics.

Variable	Terrorists (N = 48)	Non-violent (N = 96)	*T*	*U (Standardized)*
Differential associations and its conditioners				
Differential associations with terrorists	0.542 (SD = 0.504)	0.219 (SD = 0.416)	3.837***	3.880***
Network size (Computed)	478.104 (SD = 214.673)	528.083 (SD = 270.561)	−1.116	.199
Network size (Missing) [1]	518.2 (SD = 273.864)	573.533 (SD = 368.005)	-.657	.392
Posts/day (Frequency)	0.555 (SD = 0.795)	0.469 (SD = 0.442)	0.696	−1.344
Duration	38.688 (SD = 20.886)	34.365 (SD = 17.685)	1.300	1.134
Definitions				
Radical ratio	0.696 (SD = 0.397)	0.578 (SD = 0.377)	1.738†	1.804†
Differential reinforcement				
Likes/post	45.001 (SD = 47.136)	44.037 (SD = 36.296)	0.136	-.687
Comments/post	7.538 (SD = 6.813)	9.110 (SD = 9.167)	−1.051	-.161
Shares/post	0.469 (SD = 0.729)	0.156 (SD = 0.326)	2.834**	3.383***
Imitation (post type)				
Text posts (%)	17.938 (SD = 23.089)	31.271 (SD = 22.089)	−3.363**	−3.907***
Shared posts (%)	32.792 (SD = 32.854)	15.271 (SD = 20.637)	3.377***	2.556*
Picture posts (%)	45.083 (SD = 33.285)	45.577 (SD = 26.517)	−0.090	-.352
Video posts (%)	4.20 (SD = .121)	8.00 (SD = .121)	−1.798†	−2.835**

***<0.001, **<0.01, *<0.05, †<0.10.

[1] 25% missing for the terrorist group and 68% missing for the non-violent radicals group.

[5] We used software from the Israeli intelligence company Terrogence. This software has been used by law enforcement in a number of countries. See www.terrogence.com.

[6] Using the same content analysis tool used to code 'radical' posts (see page 17).

[7] Each post was coded by two coders. Due to the highly contextual nature of some posts, the coders met to reconcile a final decision where initial disagreement existed.

[8] Comments were manually examined to filter out comments made by the users themselves from the calculation.

[9] We note that there was a concern that our data contained some contamination due to timing of posts. We were concerned that some of the shares received on their final posts were made post-attack. To account for this, three different outlier tests were performed and overlapping outliers were removed. We subsequently treated these observations as missing data and used multiple imputation to fill these missing data points. Only 4 profiles included outliers of this nature.

For the conditioner factors, we coded *frequency* as the average number of posts per day made by each of the profiles. Following Smith, Wakeford, et al. (2020) *Duration* was calculated as the number of days from when the profile had been created until the last day for data collection. Following Vissers and Stolle (2014), *Network size* was first calculated as the number of friends in the friends list and converted to an 11-point scale ranging from (0) = 0–99 friends (0) up to (11) = over 1000 friends[10].

2.3.1. Descriptive statistics

Significant differences were found between the groups for a number of key variables. With regard to differential associations, terrorists had more than twice the number of posts highlighting or memorializing the terrorist attacks of their friends. The differences between the groups in the ratio of radical to non-radical posts were only marginally significant. Terrorists received a significantly larger number of shares per post, but comments and likes were nearly identical. The most common form of post was uploaded images, however terrorists had a significantly greater proportion of shared posts, and the non-violent radicals had a significantly greater proportion of originally authored text-based posts. No statistically significant differences were found for any of the conditioners of frequency, duration, or network size. There were few statistically significant inter-correlations between the variables (Appendix 1).

2.4. Analytic procedure

Our matched case-control design serves to control for individual and ecological level covariates. It is important to take this design into account, since treating this type of data as if it were a random sample can lead to poor fit and underestimated variances (Neuhaus, 1992). We employed the standard approach for matched case-control designs, the Conditional-likelihood (CL) model (Breslow, Day, Halvorsen, Prentice, & Sabai, 1978; Hosmer Lemeshow and Sturdivant, 2013). As a retrospective study with fixed status but random covariates, we use robust standard errors (Fay, Graubard, Freedman, & Midthune, 1998, pp. 195–208).

While the CL model should produce the most precise estimates for the effects of the predictors (Brookmeyer, Liang, & Linet, 1986), it cannot help in addressing our second question about classification. This is because conditional likelihoods are not equivalent to probabilities used for classification tables. As such, in order to identify how the strength of the variables in classification, we use binomial logistic regression. However, in our data, while there is independence between the cases, there is a lack of dependence within the matched pairs, which could lead to correlated error terms, and biased standard errors and tests for statistical significance (McClendon, 1994). As such, we used robust standard errors clustered on the matched pairs (Phillips et al., 2007).

3. Results

3.1. Multivariate analysis

The results of the multivariate analysis are reflective of the differences found between the groups in Table 1. In the conditional-likelihood model (Model I), terrorists had odds greater than 4 times for posting about a peer's prior terror attack (differential associations). With regard to definitions, with each additional proportion of overall posts being classified as radical, the odds of classification as a terrorists increased by 3.628 times. Terrorists were found to have produce fewer text-based posts and more shared posts by odds of 1.73 times. With regards to differential reinforcement, terrorists were more than 4.7 times more likely to receive an additional share per post, and 1.016 times more likely to receive an additional like per post (Table 3).

In the full logistic regression model (Model II), the results were highly similar. The only exception was in the effect for differential associations, where the odds of posting about a peer's prior terror were considerably larger at over 5 times. In Model III, we removed all factors not found to be statistically significant. The results remained stable, with the exception of likes per post falling below the 0.05 level, which was removed in model IV, which had all other factors remaining stable (Table 3).

3.2. Classification

Classification for model II was found to perform quite well, with an AUC of 0.8563, and an overall correct classification of 78.47%. Given that if we were to assume that all the cases were non-violent radicals we would be correct over 66% of the time, this represents a 12% increase in correct classification, which is a more than 25 percent increase in prediction of the outcome over the base rate (12/44). Given the rarity of terrorists among pools of non-violent radicals, this classification rate represents a significant improvement. For model III, the AUC was 0.8411 and classification was also 77.78%. For model IV, the AUC was 0.8379. As such, the removal of the non-statistically significant variables only affected classification by less than 1% (Table 4).

To assess sensitivity to other specifications, we raised the cutoff threshold for Model II to 0.75. While the classification rate fell slightly to 77.78%, false-positives fell to 2.08%, with only two non-violent radicals from the control group being miss-classified as cases from the terrorist group. Additionally, using an iterative approach, we found the best performing model was one that included all variables with the exception of image-based posts. This model, Model 2a, had an AUC of 0.8533, and a classification rate of 80.56%.

Table 3
Conditional-likelihood logistic regression (CL) and Binomial Logistic regression (LR) model predicting membership in the terrorist group.

Variable	Model I	Model II	Model III	Model IV
Differential associations	4.064	5.025	5.345	5.385
	(2.750)*	(2.445)**	(2.33)***	(3.44)***
Radical ratio	3.628	3.723	3.96	3.874
	(2.168)*	(2.298)*	(2.23)*	(2.15)*
Text/Shared posts	.269	.3828	.357	.348
	(.116)**	(.145)*	(.113)***	(.111)**
Video posts	.713	.729	.766	.766
	(.081)**	(.0607)***	(.065)**	(.071)**
Image posts	.991	.991	–	–
	(.011)	(.014)		
Shares/post	4.783	4.246	4.193	4.46
	(2.576)**	(1.781)***	(1.73)***	(1.83)***
Likes/post	1.016	1.014	1.00	–
	(.007)*	(.007)*	(.006)	
Comments/post	.944	.946	–	–
	(.036)	(.035)		
Frequency	2.174	1.71	–	–
	(1.262)	(.906)		
Duration	1.000	1.000	–	–
	(.000)	(.0003)		
Network size	.8378	.8859	–	–
	(.141)	(.117)		
Pseudo R^2	.5102	.3248	.2939	.2914

Model I, conditional-likelihood model, Model II-IV, logistic regression models. All estimates are Odds Ratios with robust standard errors in brackets. *** < 0.001, ** < 0.01, * < 0.05, †<0.10.

[10] Given missing data for 25% of the terrorist cases and 32% of the non-violent extremist cases, Multiple Imputation (MCMC) was used to impute the missing data points. This approach has been used in recent work comparing violent and non-violent extremists (e.g. LaFree, Jensen, James, & Safer-Lichtenstein, 2018). Following imputation, the differences between the groups remained stable (See Table 1).

Table 4
Classification statistics for models 2–4.

Model	AUC	Classification	False-positives	False-negatives
Model 2	.8563	78.47%	11.46%	41.67%
Model 2a	.8533	80.56%	8.33%	41.67%
Model 3	.8411	77.78%	11.46%	43.75%
Model 4	.8379	77.78%	11.46%	43.75%

4. Discussion

The primary objective of this study was to explore how non-text based social-media level metrics can serve to differentiate between non-violent radicals and terrorists. We analyzed theoretically-driven Facebook user and profile level behaviors and characteristics from a sample of terrorists and a matched sample of non-violent radicals. We found that each of the primary elements of social learning theory served as statistically significant predictors in classifying between the terrorist cases and non-violent radical control cases. As a retrospective case-control study that analyzed behaviors in the 100 days leading up to the attacks committed by our terrorist cases, our results have important implications for approaches to identifying potential terrorists based on social-media analysis. Our results also provide an important substantive contribution.

First and foremost, we found that within the observation period, terrorists were significantly more likely than non-violent radicals to post about a terror attack already committed by a Facebook friend. That is, terrorists often posted about attacks committed by their Facebook friends in the months prior to carrying out their own attacks. Online, terrorists, and especially lone acting ones, are often connected to each other through various direct and indirect ties (Klausen, Campion, Needle, Nguyen, & Libretti, 2016). When an individual carries out an attack, a small percentage of their network may post about, or memorialize them and their actions. Our findings suggest that future terrorists may be found among those network members who make such postings. In practice, this finding may serve as a criterion for further reducing the pool of the population whose social media accounts are surveilled, thereby increasing predictive accuracy in discriminating between non-violent radicals and potential terrorists.

One of the key premises of the social learning framework is that the likelihood of criminal behavior increases when the balance of definitions justifying the behavior outweigh those opposing it. When a known radical's postings increasingly relate to a radical ideology, cause, or group, it is indicative of fixation. This type of fixation can be a key warning signal that an individual known to hold radical attitudes may be moving towards engaging in radical behaviors, such as terrorism (Reid Meloy et al., 2012). Indeed, we found that in the 100 day observation period, the terrorist cases had a greater ratio of radical to innocuous posts, reflecting that they had a greater fixation than the non-violent radicals.

Another key finding of this study concerned the types of posts made by both the terrorist and non-violent radical cases. The majority of posts made by both groups were the uploading of images, which is in line with global trends in Facebook usage. However, the average proportion of text-based posts and shared posts were diametrically different for the two groups. While text based posts made up about 31% of the non-violent radicals' posts, they made up only 18% of the terrorists' posts. Conversely, shared posts accounted for almost 33% of the terrorists' posts, and only 15% of the non-violent radicals' posts. The ratio between originally authored text-based posts and shared posts was a statistically significant predictor is classifying between the non-violent radical and terrorist cases.

The theoretical relationship between the types of posting may be related to the social learning component of imitation. It has previously been suggested that sharing on Facebook is a form of imitation. Shares are seen to be indicative of the user's commitment to the content, or that the user feels that the content accurately reflects their opinions or feelings. As such, sharing of content is a way of expression that requires less cognitive sophistication and ability than writing (Kaur, Balakrishnan, Rana, & Sinniah, 2019). In this regard, Baele (2017) found that in their writings, terrorists displayed lower levels of cognitive sophistication than non-violent radicals. There is also a growing group of scholars who suggest that radical social media activity can act as a protective factor against violent expressions, by providing a non-violent outlet to voice grievances (e.g. Kardaş & Özdemir, 2018; Taylor, Holbrook, & Joinson, 2017). Our results may lend support to this argument.

Perhaps more importantly however, these results further highlight the potential limitations of text-based analysis. According to the results of this study, text based analysis would have only been able to analyze some 31% of the content produced by the non-violent radicals, and only slightly more than half of that for the terrorist cases. The difficulties for text-based analysis to account for the majority of content being produced by a significantly larger non-violent radical population (Shortland, 2016) is now further compounded by the fact that non-violent radicals may also be producing almost double the volume of written material of terrorists. This finding further strengthens the argument for combining behavioral elements into detection models, either in lieu or support of text-based analysis.

With respect to the final element of social learning theory, differential reinforcement, we found that receiving more likes and shares per post, but not comments, were significant predictors for being classified as a terrorist over a non-violent radical. The small effect size for likes per post reflects the fact that likes are the most prevalent type of response. According to models 1 and 2, for each additional like received per post, the odds of being classified as a terrorist case increased between 1.4 and 1.6%. The much larger effect size for shares per post reflects the fact that this type of impression is the most infrequent type of interaction. While the receiving of likes has been shown to increase social support and connectedness (Wohn, Carr, & Hayes, 2016; Zell & Moeller, 2018), receiving more shares increases sense of influence and is indicative of the receiver being viewed as an opinion leaders in their network. Opinion leaders may be pressured to conform to the type of behavior that is expected of them (Oeldorf-Hirsch & Sundar, 2015). That terrorists received more shares per post may also relate to their proclivity for sharing of other posts as shared pieces are more likely to be re-shared than originally-authored posts (Guerini, Staiano, & Albanese, 2013). Additionally, radical posts may generate more shares than non-radical posts, since more negative, emotional and harrowing posts also tend to garner more shares (Harber & Cohen, 2005; Harber, Podolski, & Dyer, 2014).

Differentiating terrorists from a large pool of non-violent radicals based on their social media activities is an exceptionally difficult task. Some have estimated that with the current technological and methodological approaches to automated detection, namely text-based analysis of large swaths of the population, there may be as many as 100,000 false positives for every terrorist (Munk, 2017). We believe that automated detection ought to be a tool to assist analysts to sort through the data of a manually selected sampling pool of known radicals. Our results support this approach and demonstrate how social media behaviors can be leveraged to more accurately discriminate terrorists from non-violent radicals in a sample of known, and seemingly similar radicals (Cohen et al., 2014; Pelzer, 2018). As discussed above, the Israeli case is even more challenging, given the high prevalence of non-violent radicals, or individuals who support and justify terrorism. But as we have demonstrated, even in such a case it is possible to achieve an acceptable classification and false-positive rate. As such, a similar approach ought to be even more successful in a situation in which the base-rate for radical attitudes is even smaller (Neuman et al., 2019). Reducing false-positives should be the goal of any security service, who should be interested in reducing the number of false-arrests and the potential for them to increases stigmatization and lead to a backlash effect (Tankebe, 2020).

Of course our study is not without its limitations. For example, we

did not examine the relationship between post type and differential reinforcement, or other possible interactions. Furthermore, our focus in this study was on one specific platform, Facebook. As such, we have not accounted for the role of other platforms. This limitation is difficult to overcome and we have not identified any other research which has examined a single sample over multiple platforms in such a way. Furthermore, our dataset was limited to a specific context, Israel, which we acknowledge is unique. At the same time, terrorism research from Israel has been shown to provide important lessons to other western contexts. We encourage future research that applies similar approaches to other contexts. Lastly, we did not examine the definitions themselves. This decision was made on account of the fact that our goal was to move beyond text-based analysis. Nevertheless, as stated above, future research should seek to integrate behavior-level metrics and text, sentiment, and semantic analysis. Despite these limitations, we believe that the metrics we identified have the potential to be generalizable. We also believe that our study provides an important, substantive contribution to the literature. In this regard, our approach highlights the usefulness of traditional criminological frameworks for informing the development of evidence-based, big-data solutions (Chan & Bennett Moses, 2016).

5. Conclusion

Since social media is used by individuals to express themselves, it holds significant potential to be leveraged as a window of opportunity. The lack of theoretically driven approaches has hindered the development of the identification of factors that distinguish the online behaviors of violent and non-violent radicals. This has had serious implications for the advancement of computational and predictive approaches to terrorism prevention, and evidence based policy and practice more generally. In this study we have identified that a number of user and profile level behaviors and characteristics displayed on Facebook profiles can differentiate between terrorists and non-violent radicals. Jointly and severely, these metrics provide promise for moving beyond exclusively on text based analysis. Given that text-based posts represent only a small proportion of overall activity, making advances in this direction is even more necessary. In fact, many of these metrics may be superior to written content even where it does exist. Or at the least could serve to greatly strengthen the predictive power and accuracy of detection tools employing text-based analysis.

It is our hope that this study will also encourage other researchers about the potential for utilizing small but rich datasets to examine aspects of the internet and radicalization. Whilst it may not be possible to identify a profile of a terrorist offender per se, specific sets of behaviors or patterns of activities may be more easily classified as indicators of the move from radical beliefs to radical behaviors. Future work should seek analyze changes in online behaviors, ideally with cross-sectional time series (panel) data.

Credit author statement

Michael Wolfowicz: Conceptualization, Formal analysis, Methodology, Writing-Reviewing and editing, Project administration Simon Perry: Conceptualization, Methodology, Writing – original draft, Project administration Badi Hasisi: Supervision, Writing – review & editing, Funding acquisition David Weisburd: Supervision Writing - Review & Editing, Funding acquisition.

Ethics statement

All data was anonymized following coding and we have not included any examples of actual posts made by individuals in the dataset. By presenting only aggregate level data, it is not possible for individuals in the dataset to be identified. Additionally, at least with respect to terrorists in our dataset, the majority, if not all of the profiles are no longer active, due to removal by either the platform, security services, or individuals associated with the profile's creator. The use of the publicly available data in this study is in line with Facebook's terms and conditions relating to accessibility and usage, including for research purposes. This study received approval by Hebrew University's Institutional review board.

Funding details

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 699824. Additional funding and assistance was received from The Federmann Cyber Security Center—Cyber Law Program at the Hebrew University of Jerusalem. This work was also made possible thanks to resources provided by Terrogence.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2020.106646.

Appendix 1Pairwise correlations between continuous variables

	Posts/day	Network size	Duration	Radical ratio	Likes/post	Comments/post	Shares/post	Video posts	Pic posts	Text/shared posts
Posts/day	1									
Network size	0.1627	1								
Duration	0.0514	−0.1183	1							
Radical ratio	−0.1103	−0.046	0.1375	1						
Likes/post	−0.161	0.1880*	−0.0207	0.0117	1					
Comments/post	−0.2323*	0.0888	0.0453	−0.0747	0.7227*	1				
Shares/post	0.1223	0.045	−0.024	0.1407	0.1720*	0.0835	1			
Video posts	−0.0983	−0.2071*	0.0839	0.2584*	−0.0012	0.0244	0.0735	1		
Pic posts	−0.2570*	−0.0304	−0.0462	−0.2037*	0.0552	0.1375	−0.1258	−0.2996*	1	
Text/shared posts	0.2150*	−0.0066	0.0421	0.0675	−0.2509*	−0.2374*	0.132	−0.0522	−0.1695*	1

*<0.05.

References

Akers, R. L. (1998). *Social learning and social structure: A general theory of crime and deviance*. Transaction Publishers.

Akers, R. L., & Sellers, C. S. (2004). *Criminological theories. 4*. Los Angeles, CA: Roxbury Publishing.

Akers, R. L., & Silverman, A. L. (2014). Toward a social learning model of violence and terrorism. In M. A. Zahn, H. H. Brownstein, & S. L. Jackson (Eds.), *Violence: From theory to research* (pp. 27–44). Routledge.

Akins, J. K., & Winfree, L. T., Jr. (2016). Social learning theory and becoming a terrorist: New challenges for a general theory. *The Handbook of the Criminology of Terrorism, 133*.

Baele, S. J. (2017). Lone-actor terrorists' emotions and cognition: An evaluation beyond stereotypes. *Political Psychology, 38*(3), 449–468.

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication, 28*(3), 12–29.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice- Hall.

Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V., & Ledovaya, Y. (2018). Dark personalities on Facebook: Harmful online behaviors and language. *Computers in Human Behavior, 78*, 151–159.

Breslow, N. E., Day, N. E., Halvorsen, K. T., Prentice, R. L., & Sabai, C. (1978). Estimation of multiple relative risk functions in matched case-control studies. *American Journal of Epidemiology, 108*(4), 299–307.

Brookmeyer, R. O. N., Liang, K. Y., & Linet, M. (1986). Matched case-control designs and overmatched analyses. *American Journal of Epidemiology, 124*(4), 693–701.

Brown, R. C., Bendig, E., Fischer, T., Goldwich, A. D., Baumeister, H., & Plener, P. L. (2019). Can acute suicidality be predicted by instagram data? Results from qualitative and quantitative language analyses. *PloS One, 14*(9).

Brynielsson, J., Horndahl, A., Johansson, F., Kaati, L., Mårtenson, C., & Svenson, P. (2013). Harvesting and analysis of weak signals for detecting lone wolf terrorists. *Security Informatics, 2*(1), 1–15.

Chan, J., & Bennett Moses, L. (2016). Is big data challenging criminology? *Theoretical Criminology, 20*(1), 21–39.

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *NPJ digital medicine, 3*(1), 1–11.

Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M., & Agrawal, P. (2019). Understanding emotions in text using deep learning and big data. *Computers in Human Behavior, 93*, 309–317.

Cohen, K., Johansson, F., Kaati, L., & Mork, J. C. (2014). Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence, 26*(1), 246–256.

Cone, H. A. (2016). *Differential Reinforcement in the online Radicalization of western muslim women converts (doctoral dissertation*. Walden University.

Costello, M., Hawdon, J., Ratliff, T., & Grantham, T. (2016). Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior, 63*, 311–320.

Curtiss, M., Becker, I., Bosman, T., Doroshenko, S., Grijincu, L., Jackson, T., & Shen, G. (2013). Unicorn: A system for searching the social graph. *Proceedings of the VLDB Endowment, 6*(11), 1150–1161.

van Dam, J. W., & Van De Velden, M. (2015). Online profiling and clustering of Facebook users. *Decision Support Systems, 70*, 60–72.

Dillon, L., Neo, L. S., & Freilich, J. D. (2019). *A comparison of ISIS foreign fighters and supporters social media posts: An exploratory mixed-method content analysis*. Behavioral Sciences of Terrorism and Political Aggression.

D'Angelo, J., Kerr, B., & Moreno, M. A. (2014). Facebook displays as predictors of binge drinking: From the virtual to the visceral. *Bulletin of Science, Technology & Society, 34*(5–6), 159–169.

Eftekhar, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Computers in Human Behavior, 37*, 162–170.

Farahbakhsh, R., Han, X., Cuevas, A., & Crespi, N. (2013). Analysis of publicly disclosed information in facebook profiles. August. In *2013 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2013)* (pp. 699–705). IEEE.

Fay, M. P., Graubard, B. I., Freedman, L. S., & Midthune, D. N. (1998). *Conditional logistic regression with sandwich estimators: Application to a meta-analysis*. Biometrics.

Ferrara, E., Wang, W. Q., Varol, O., Flammini, A., & Galstyan, A. (2016). November). Predicting online extremism, content adopters, and interaction reciprocity. In *International conference on social informatics* (pp. 22–39). Cham: Springer.

Freilich, J. D., & LaFree, G. (2016). Measurement issues in the study of terrorism: Introducing the special issue. *Studies in Conflict and Terrorism, 39*(7–8), 569–579.

Frissen, T. (2020). Internet, the great radicalizer? Exploring relationships between seeking for online extremist materials and cognitive radicalization in young adults. *Computers in Human Behavior*, 106549.

Frost, R. L., & Rickwood, D. J. (2017). A systematic review of the mental health outcomes associated with Facebook use. *Computers in Human Behavior, 76*, 576–600.

Gaspar, R., Pedro, C., Panagiotopoulos, P., & Seibt, B. (2016). Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior, 56*, 179–191.

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist use of the Internet by the numbers: Quantifying behaviors, patterns, and processes. *Criminology & Public Policy, 16*(1), 99–117.

Gill, P., Horgan, J., Corner, E., & Silver, J. (2016). Indicators of lone actor violent events: The problems of low base rates and long observational periods. *Journal of Threat Assessment and Management, 3*(3–4), 165.

Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. (2010). March). Walking in facebook: A case study of unbiased sampling of osns. In *2010 proceedings IEEE infocom* (pp. 1–9) (Ieee).

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking, 14*(9), 483–488.

Guerini, M., Staiano, J., & Albanese, D. (2013). Exploring image virality in google plus. September. In *2013 international conference on social computing* (pp. 671–678). IEEE.

Harber, K. D., & Cohen, D. J. (2005). The emotional broadcaster theory of social sharing. *Journal of Language and Social Psychology, 24*(4), 382–400.

Harber, K. D., Podolski, P., & Dyer, L. (2014). Hearing stories that violate expectations leads to emotional broadcasting. *Journal of Language and Social Psychology, 33*(1), 5–28.

Hasisi, B., Carmel, T., Weisburd, D., & Wolfowicz, M. (2019). Crime and terror: Examining criminal risk factors for terrorist recidivism. *Journal of Quantitative Criminology*, 1–24.

Hasisi, B., Perry, S., & Wolfowicz, M. (2019). Counter-terrorism effectiveness and human rights in Israel. *International human rights and counter-terrorism*, 409–429.

Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter? *American Journal of Sociology, 106*(4), 1013–1057.

Haynie, D. L. (2002). Friendship networks and delinquency: The relative nature of peer delinquency. *Journal of Quantitative Criminology, 18*(2), 99–134.

Haynie, D. L., Doogan, N. J., & Soller, B. (2014). Gender, friendship networks, and delinquency: A dynamic network approach. *Criminology, 52*(4), 688–722.

Holbrook, D., & Taylor, M. (2014). Developing grading processes for ideological content. *Journal of Policing, Intelligence and Counter Terrorism, 9*(1), 32–47.

Holt, T. J., & Bossler, A. M. (2016). *Cybercrime in progress: Theory and prevention of technology-enabled offenses*. Routledge.

Holt, T. J., Burruss, G. W., & Bossler, A. M. (2010). Social learning and cyber-deviance: Examining the importance of a full social learning model in the virtual world. *Journal of Crime and Justice, 33*(2), 31–61.

Hong, E., & Whitehead, L. (2014). Facebook groups: Perception and usage among undergraduates in the context of learning. *International conference on information systems*.

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hussain, J., Satti, F. A., Afzal, M., Khan, W. A., Bilal, H. S. M., Ansaar, M. Z., & Park, G. H. (2019). Exploring the dominant features of social media for depression detection. *Journal of Information Science*, Article 0165551519860469.

Kaati, L., Shrestha, A., & Cohen, K. (2016). June). Linguistic analysis of lone offender manifestos. In *2016 IEEE international conference on cybercrime and computer forensic (ICCCF)* (pp. 1–8). IEEE.

Kardaş, T., & Özdemir, Ö. B. (2018). The making of European foreign fighters: Identity, social media and virtual radicalization. In *Non-state armed actors in the Middle East* (pp. 213–235). Cham: Palgrave Macmillan.

Kaur, W., Balakrishnan, V., Rana, O., & Sinniah, A. (2019). Liking, sharing, commenting and reacting on Facebook: User behaviors' impact on sentiment intensity. *Telematics and Informatics, 39*, 25–36.

Kim, C., & Yang, S. U. (2017). Like, comment, and share on Facebook: How each behavior differs from the other. *Public Relations Review, 43*(2), 441–449.

Kingston, D. A., Fedoroff, P., Firestone, P., Curry, S., & Bradford, J. M. (2008). Pornography use and sexual aggression: The impact of frequency and type of pornography use on recidivism among sexual offenders. *Aggressive Behavior, 34*(4), 341–351.

Klausen, J., Campion, S., Needle, N., Nguyen, G., & Libretti, R. (2016). Toward a behavioral model of "homegrown" radicalization trajectories. *Studies in Conflict & Terrorism, 39*(1), 67–83.

Klausen, J., Libretti, R., Hung, B. W., & Jayasumana, A. P. (2018). Radicalization trajectories: An evidence-based computational approach to dynamic risk assessment of "homegrown" jihadists. *Studies in Conflict & Terrorism*, 1–28.

Kuo, C. L., Duan, Y., & Grady, J. (2018). Unconditional or conditional logistic regression model for age-matched case–control data? *Frontiers in public health, 6*, 57.

Kurant, M., Markopoulou, A., & Thiran, P. (2010). On the bias of BFS (breadth first search). September. In *2010 22nd international teletraffic congress (lTC 22)* (pp. 1–8). IEEE.

LaFree, G., Jensen, M. A., James, P. A., & Safer-Lichtenstein, A. (2018). Correlates of violent political extremism in the United States. *Criminology, 56*(2), 233–268.

Li, Z., Fan, Y., Jiang, B., Lei, T., & Liu, W. (2019). A survey on sentiment analysis and opinion mining for social multimedia. *Multimedia Tools and Applications, 78*(6), 6939–6967.

McClendon, M. J. (1994). *Multiple regression and causal analysis*. Itasca, IL. *IL*: FE Peacock Publishers.

McCuddy, T., & Vogel, M. (2015a). Beyond traditional interaction: Exploring the functional form of the exposure-offending association across online network size. *Journal of Criminal Justice, 43*(2), 89–98.

McCuddy, T., & Vogel, M. (2015b). More than just friends: Online social networks and offending. *Criminal Justice Review, 40*(2), 169–189.

Miller, B., & Morris, R. G. (2016). Virtual peer effects in social learning theory. *Crime & Delinquency, 62*(12), 1543–1569.

Minkus, T., Ding, Y., Dey, R., & Ross, K. W. (2015). November). The city privacy attack: Combining social media and public records for detailed profiles of adults and children. In *Proceedings of the 2015 ACM on conference on online social networks* (pp. 71–81). ACM.

Munk, T. B. (2017). 100,000 false positives for every real terrorist: Why anti-terror algorithms don't work. *First Monday, 22*(9).

Ness, A. M., Johnson, G., Ault, M. K., Taylor, W. D., Griffith, J. A., Connelly, S., & Jensen, M. L. (2017). Reactions to ideological websites: The impact of emotional appeals, credibility, and pre-existing attitudes. *Computers in Human Behavior, 72*, 496–511.

Neuhaus, J. M. (1992). Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research, 1*(3), 249–273.

Neuman, Y., Cohen, Y., & Neuman, Y. (2019). How to (better) find a perpetrator in a haystack. *Journal of Big Data, 6*(1), 9.

Nouh, M., Nurse, R. J., & Goldsmith, M. (2019). Understanding the radical mind: Identifying signals to detect extremist content on twitter. July. In *2019 IEEE international conference on intelligence and security informatics (ISI)* (pp. 98–103). IEEE.

Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior, 44*, 240–249.

Ophir, Y., Asterhan, C. S., & Schwarz, B. B. (2019). The digital footprints of adolescent depression, social rejection and victimization of bullying on Facebook. *Computers in Human Behavior, 91*, 62–71.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior, 31*, 527–541.

Parekh, D., Amarasingam, A., Dawson, L., & Ruths, D. (2018). Studying jihadists on social media: A critique of data collection methodologies. *Perspectives on Terrorism, 12*(3), 5–23.

Patton, D. U., Hong, J. S., Ranney, M., Patel, S., Kelley, C., Eschmann, R., et al. (2014). Social media as a vector for youth violence: A review of the literature. *Computers in Human Behavior, 35*, 548–553.

Pauwels, L., & Schils, N. (2016). Differential online exposure to extremist content and political violence: Testing the relative strength of social learning and competing perspectives. *Terrorism and Political Violence, 28*(1), 1–29.

Pelzer, R. (2018). Policing of terrorism using data from social media. *European Journal of Scientific Research, 3*(2), 163–179.

Peterson, J., & Densley, J. (2017). Cyber violence: What do we know and where do we go from here? *Aggression and Violent Behavior, 34*, 193–200.

Phillips, S., Matusko, J., & Tomasovic, E. (2007). Reconsidering the relationship between alcohol and lethal violence. *Journal of Interpersonal Violence, 22*(1), 66–84.

Pratt, T. C., Cullen, F. T., Sellers, C. S., Thomas Winfree, L., Jr., Madensen, T. D., Daigle, L. E., … Gau, J. M. (2010). The empirical status of social learning theory: A meta-analysis. *Justice Quarterly, 27*(6), 765–802.

Pressman, D. E., & Ivan, C. (2016). Internet use and violent extremism: A cyber-vera risk assessment protocol. In M. Khader, L. S. Neo, G. Ong, E. T. Mingyi, & J. Chin (Eds.), *Combating violent extremism and radicalisation in the digital era* (pp. 391–409). Hershey, PA: IGI Global Publishers.

Quiggin, T. (2017). On and off the radar: Tactical and strategic responses to screening known potential terrorist attackers. *Perspectives on terrorism, 11*(5), 42–50.

Reid Meloy, J., Hoffmann, J., Guldimann, A., & James, D. (2012). The role of warning behaviors in threat assessment: An exploration and suggested typology. *Behavioral Sciences & the Law, 30*(3), 256–279.

Schmid, A. P., & Forest, J. J. (2018). Research desiderata: 150 un-and under-researched topics and themes in the field of (counter-) terrorism studies–a new list. *Perspectives on terrorism, 12*(4), 68–76.

Scrivens, R., Davies, G., & Frank, R. (2018). Searching for signs of extremism on the web: An introduction to sentiment-based identification of radical authors. *Behavioral Sciences of Terrorism and Political Aggression, 10*(1), 39–59.

Scrivens, R., Davies, G., & Frank, R. (2020). Measuring the evolution of radical right-wing posting behaviors online. *Deviant Behavior, 41*(2), 216–232.

Scrivens, R., Gaudette, T., Davies, G., & Frank, R. (2019a). *Searching for extremist content online using the dark crawler and sentiment analysis. Methods of Criminology and Criminal Justice Research* (Vol. 24, pp. 179–194). Sociology of Crime, Law and Deviance. Emerald Publishing Limited.

Scrivens, R., Gill, P., & Conway, M. (2019b). The role of the internet in facilitating violent extremism and terrorism: Suggestions for progressing research. In *The palgrave handbook of international cybercrime and cyberdeviance*. London, UK: Palgrave (Cybercrime Series)*(by invitation). Forthcoming, 2*.

Seng, N. L., Khader, M., & Pang, J. S. (2018). *Comparing ISIS foreign fighters versus sympathisers: Insights from their twitter postings*. HOME TEAM.

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking, 21*(4), 217–228.

Shadmanfaat, S. M., Howell, C. J., Muniz, C. N., Cochran, J. K., Kabiri, S., & Fontaine, E. M. (2020). Cyberbullying perpetration: An empirical test of social learning theory in Iran. *Deviant Behavior, 41*(3), 278–293.

Shapiro, L. R., & Maras, M. H. (2019). Women's radicalization to religious terrorism: An examination of ISIS cases in the United States. *Studies in Conflict & Terrorism, 42*(1–2), 88–119.

Shortland, N. D. (2016). "On the internet, nobody knows you're a dog": The online risk assessment of violent extremists. In M. Khader (Ed.), *Combating violent extremism and radicalization in the digital era* (pp. 349–373). Hershey PA: IGI Global.

Skinner, W. F., & Fream, A. M. (1997). A social learning theory analysis of computer crime among college students. *Journal of Research in Crime and Delinquency, 34*(4), 495–518.

Smith, L. G., Blackwood, L., & Thomas, E. F. (2020). The need to refocus on the group as the site of radicalization. *Perspectives on Psychological Science, 15*(2), 327–352.

Smith, L. G., Wakeford, L., Cribbin, T. F., Barnett, J., & Hou, W. K. (2020). Detecting psychological change through mobilizing interactions and changes in extremist linguistic style. *Computers in Human Behavior, 108*, 106298.

Sutch, H., & Carter, P. (2019). Anonymity, membership-length and postage frequency as predictors of extremist language and behaviour among twitter users. *International Journal of Cyber Criminology, 13*(2), 439–459.

Tankebe, J. (2020). Unintended negative outcomes of counter-terrorism policing: Procedural (in) justice and perceived risk of recruitment into terrorism. In *Understanding recruitment to organized crime and terrorism* (pp. 105–119). Cham: Springer.

Taylor, P. J., Holbrook, D., & Joinson, A. (2017). A same kind of different: Affordances, terrorism and the internet. *Criminology & Public Policy, 16*(1), 127–133.

Vissers, S., & Stolle, D. (2014). Spill-over effects between facebook and on/offline political participation? Evidence from a two-wave panel study. *Journal of Information Technology & Politics, 11*(3), 259–275.

Wohn, D. Y., Carr, C. T., & Hayes, R. A. (2016). How affective is a "Like"?: The effect of paralinguistic digital affordances on perceived social support. *Cyberpsychology, Behavior, and Social Networking, 19*(9), 562–566.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of Quantitative Criminology, 36*(3), 407–447.