Investigating Non-suicidal Self-injury Discussions on Twitter

Muhammad Abubakar Alhassan, and Diane Pennington.

Abstract—Social networking sites have become a space for people to discuss public health issues such as non-suicidal self-injury (NSSI). There are thousands of tweets containing self-harm and self-injury hashtags on Twitter. It is difficult to distinguish between different users who participate in self-injury discussions on Twitter and how their opinions change over time. Also, it is challenging to understand the topics surrounding NSSI discussions on Twitter. We retrieved tweets using *#selfham* and *#selfinjury* hashtags and investigated those from the United kingdom. We applied inductive coding and grouped tweeters into different categories. This study used the Latent Dirichlet Allocation (LDA) algorithm to infer the optimum number of topics that describes our corpus. Our findings revealed that many of those participating in NSSI discussions are non-professional users as opposed to medical experts and academics. Support organisations, medical teams, and academics were campaigning positively on raising self-injury awareness and recovery. Using LDAvis visualisation technique, we selected the top 20 most relevant terms from each topic and interpreted the topics as; *children and youth well-being*, *self-harm misjudgement*, *mental health awareness*, *school and mental health support* and, *suicide and mental-health issues*. More than 50% of these topics were discussed in England compared to Scotland, Wales, Ireland and Northern Ireland. Our findings highlight the advantages of using the Twitter social network in tackling the problem of self-injury through awareness. There is a need to study the potential risks associated with the use of social networks among self-injurers.

Keywords—self-harm, non-suicidal self-injury, Twitter, social networks.

I. Introduction

To date, there is no globally recognised standard definition of intentional self-injury. The term is usually recognised as Deliberate Self-Harm (DSH) in the united kingdom and Non-Suicidal Self-Injury (NSSI) in the United States of America, Australia and Canada. Experts have linked the behaviour to burning or cutting the body (skin) with a razor or a knife to relieve stress, anger, or frustration. Some researchers described self-injury as an act of hurting oneself to overcome the emotional pain that a person suffers as a result of abuse, rape, bullying and many other factors. Although hurting oneself does not relieve anxiety or complicated feelings, some self-injurers believe that hurting themselves is a solution to overcoming emotional distress. It works for only a short time. Engaging in this behaviour becomes an addiction for them, and the urges are uncontrollable.

Although there are several reasons people engage in self-inflicted behaviours, some practice the behaviour for the religious (Tatbir by Shia) purposes and not for the sake of relieving their anxiety or resentment [1]. Also, others do it to get rid of stress. Self-injury is well-known in our communities, and research has found that this behaviour is prevalent among young people, especially girls. Nowadays, many young individuals are the most active users of social media applications. These tools are among the most popular ways for people to share information about their lives and mental well-being. While self-harming individuals hide their behaviour, social media has become a platform for them to come out and express their concerns and even seek help. They can disclose their difficult feelings on social media sites because they can keep their privacy anonymously. Doing this allows them to connect with online users and build relationships in a social passion.

On Twitter, users can interact with online community members by composing and posting a short text (tweet). Self-injurers tend to look for help online, and Twitter is one of the numerous online spaces they engage in building social connections. The platform receives thousands of tweets posted by users that contained non-suicidal self-injury hashtags that need to be categorized using current state of the art machine learning (unsupervised) algorithms. Also, there is insufficient research on what users are discussing twitter regarding deliberate self-harm. Even though there is an increasing concern in this research area, our knowledge of people tweeting about self-harm on Twitter and their opinions regarding this public health issue is inadequate. Furthermore, we have little or no understanding of the impacts of tweets containing NSSI hashtags on Twitter. There could be a reciprocation between the positive and negative effects such as; creating an online social community to avoid social isolation and exposure to digital self-harm tweets that encourage the behaviour.

II. Related work

1) Self-harm and social networking sites: Twitter is one of the largest social networking sites, with over 100 million users around the globe. The platform allows users to post information by writing tweets in a text that is not more than 280 characters in length. Although users can reply to tweets, they can also share or retweet information in a social passion supported by the platform. However, Twitter is among the social networking sites where people discuss a wide range of topics, including their mental health issues. Although many researchers in mental health and social media focus on investigating self-injurious contents on social media, most of their research focuses on NSSI images and videos that people post or share on social media with less attention on NSSI textual contents. Many of these investigations were carried out on Instagram, youtube, and Flickr platforms.

Muhammad and Diane are with the Department of Computer and Information Sciences, University of Strathclyde, Glasgow, G1 1XH, United Kingdom. E-mails: muhammad.alhassan@strath.ac.uk diane.pennington@strath.ac.uk

Every day, people use hashtags to create and post information on social media like Twitter and Instagram. This information often covered many hidden insights that self-injurers are sharing on these platforms. A recent investigation discovered some ambiguous hashtags that self-injurers are using on Instagram to communicate about NSSI [2]. Their findings demonstrate the use of "#Mysecretfamily" hashtag as an umbrella tag containing *"#cat"* and *"#deb"* that members of the online community are using in posting graphic information about cutting and depression. They further emphasized that the caution warning message of these graphic contents are not reliable.

Another comparative study examined NSSI contents on Tumblr, Twitter, and Instagram [3]. The researchers searched these platforms using *"#cutting"* hashtag and retrieved a set of relevant contents for their investigations. Moreover, the posts they examined do not encourage users to seek either formal or informal support. They discovered that the percentage of posts containing NSSI graphic contents on Instagram outweighs that of Twitter. This result is similar to the previous study findings that investigated NSSI images on Instagram [4]. The findings of [3] were also confirmed by a recent investigation that examined the three platforms and discovered how users share NSSI images [5]. More than 50% of the pictures they examined have no explicit NSSI contents. In some images with explicit contents, it was discovered that the most common explicit representation is self-injury [5]. Their research proved that none of the analysed images encourages intentional self-injury and themes such as identity and belonging, addiction and recovery were discovered.

Furthermore, in one of the pioneer studies that analyzed NSSI contents on Instagram, it was discovered that images depicting cutting or NSSI wounds attract more comments than non-graphic photos [4]. Similarly, it was found that exposure to self-injurious behaviour on Instagram is associated with self-harm, emotional difficulties, and suicidal thoughts [6]. Similarly, some researchers significant variations in terms of linguistic, readability and sentiments between NSSI and non-NSSI contents on Flickr [7]. The researchers combined these contents' characteristics and features and proposed an automatic framework (using both supervised and non-supervised approaches) that predicts NSSI contents. On the other hand, some existing studies validate NSSI video contents on social media with more emphasis on YouTube. This platform is one of the world-leading social media applications that allow users from different places to search and view NSSI videos which could be helpful or harmful to the audiences [8]. Viewers and video presenters often interact with one another through exchanging comments on a video. Most of the NSSI videos' responses were found to be directed towards self-disclosure [8]. Despite the fact some of the audiences were reporting that they are self-harming, only a few comments from the NSSI videos examine by [8] were advising for self-recovery.

As NSSI often involves cutting of the skin, there is speculation that users share first aid tips or ways to treat injuries without necessarily attending medical treatments. Some researchers scrutinise the content of YouTube videos discussing first aid for people that are self-harming [9]. Their findings demonstrate that these videos could reinforce NSSI, and there is a tendency that viewers could not seek medical help regarding NSSI. Also, some authors qualitatively examined the effect of blue whale challenge game on Twitter and YouTube as the game reinforces self-harm and eventually encourages players that loose the game to end their lives [10]. They discovered that users participated in the game to raise awareness and advise on not playing the game. On the other hand, around 50% of the YouTube videos they examined failed to comply with standard guidelines of the Suicide Prevention Resource Center.

There are limited investigations on NSSI contents ton Twitter. The work of [11] explored a set of tweets posted using "#cutting" on Twitter. Although the investigator founds themes such as; celebrity influence, self-harm is not a joke; her study demonstrates how individuals misunderstand self-harm on a Twitter social network. The hashtag used by [11] could not capture many relevant tweets surrounding self-harm discussions on Twitter because some people used the tag to post irrelevant tweets related to haircutting style. The approach used by [11] could be tedious and time consuming for a large volume of tweets discussing NSSI on Twitter.

A. User profile studies

A deep understanding of Twitter users involves a proper investigation of the account handle through examining the reported information users provided in their profile which is often called the user bio-descriptions. Many tweeters provide information about their locations, occupations, and personal interest in their account description. Existing studies leveraged this information and study Twitter users in different fields. For example, [12] analysed the gender, geographic locations, and race of tweeters from the United States. Similarly, [13] studied Twitter users' demographic information in the UK and proposed a tool that automatically detects the social class, occupation, and age of the users. Similar research was conducted by [14] in which the authors analysed the use of language (in large scales) on Twitter and detected the occupational class of users in the UK. Although some of the existing studies focused on using machine learning techniques to infer a tweeter's occupational class or age, another study used content analysis. The authors examined Twitter user's groups ranging from individuals to corporate organisations, and news media handles [15], [16].

B. Emotions and NSSI

Most researchers believe there is a link between self-harm behaviour and negative emotions [17]. It was learned that there is a significant change in daily emotions between self-injurers and individuals that are not self-harming. People who are suffering from self-harm tend to experience more negative emotions than those who do not engage in NSSI [17]. Self-injurers experience difficult emotions as a result of traumatic experiences such as bullying, sexual or child abuse [18]. These and many other factors make them consider self-harm behaviour as a coping mechanism. Many of the existing studies on self-harm behaviours and emotions were based on survey approaches [6]. There has been little or less attention to how

Twitter users express their opinions to self-injurers through posting tweets containing self-harm and self-injury hashtags. Positive attitudes and opinions towards self-injurers play a significant role in achieving effective recovery.

C. Topic Modelling

One of the existing current state-of-the-art statistical topic models is the Latent Dirichlet Allocation (LDA) proposed by [19]. The model view documents as a collection of multiple topics offer a powerful and unique way of identifying unknown themes in large document collections. In the LDA model, topics or themes of a text corpus are a set of words in a document and the distribution of these words provides room for observations and inferences [19]. This model is crucial in analysing large textual data that is in unstructured form. Many existing research studies applied this model in different investigations, such as climate change discussions by [20]. The investigators used LDA and examined themes centred around changed in climate from different nations. Also, [21] applied LDA and explored tweeters discussions on economic challenges at the time of the 2012 US presidential elections.

On the other hand, tweets related to mental health issues were also investigated by [15]. The researchers analysed tweets related to dementia conversations on Twitter. These studies and many other investigations demonstrate the importance and effectiveness of using LDA to understand the topics individuals discuss on Twitter. Inspired with the LDA algorithm's robust nature, our study considered using the algorithm to understand the themes surrounding self-injury discussions on Twitter.

However, considering the research studies conducted by many authors in this field, Our knowledge of what people are discussing regarding NSSI is unclear. Although social networking sites could be a speaker through which the voices of self-injurers could be heard, we lack understanding of the online individuals and communities that are reaching out to self-harming users. Our research expands on the work of [11] and answered the following research questions.

D. Research Questions

- Who is tweeting NSSI on Twitter? In this question, we aim to categorise people who use *#selharm* and *#selfinjury* hashtags and posts information about NSSI on Twitter. Addressing this question will broader our understanding of different group of users involved in self-injury discussions on Twitter.
- What tweeters tweet about self-injury on Twitter?
  In this question, we aim to understand the main topics that users tweet about intentional self-injury from diverse locations. This is to understand the impact of tweeting non-suicidal self-injury on Twitter by using current state-of-the-art topic modelling algorithms.
- What is the opinion of users tweeting self-injury on Twitter? We are interested in understanding the user's sentiments in tweeting about NSSI on Twitter in this question. Knowing their opinions will shed more light on how online users influence or understand the importance of self-injury. This will increase our understanding of the benefits or risks associated with social networks' for people with self-injury.

E. Why Twitter?

Although modern social media technology offers data across various platforms, our study focused on the Twitter social network. It is among the leading social networks, with more than 300 million users producing around 500 million tweets a day. In this study, we purposely focused on Twitter because of two reasons. Firstly, Twitter serves as a data source for many researchers and is believed to impact mental health informatics [22] significantly. Secondly, the platform provides useful meta-data information that can be used to understand users, their opinion and level of social engagements. For example, even though tweets are short texts, they contained some key attributes such as the user who created them, the time they were posted on Twitter, and the language used in composing the tweet. During our data analysis, all these information were taken into account as they provide more in-depth descriptions of the data.

III. Methods

Our study retrieved data from Twitter social media using Twitter Archive Google Sheets (TAGS). It is a tool proposed by Martin Hawksey, who is an educational technologist [23]. We purposely used TAGS as the tool allows for cloud storage using Google excel sheets and provides an easy way of accessing archived tweets at any given time using multiple devices. Although accessing TAGS requires authentication through Twitter user or developer account, the tool effectively enables users to build a personalised spreadsheet database collection of tweets based on a specific query term or hashtag in compliance with Twitter's conditions.

Our research searched for tweets discussing NSSI on Twitter using *#selfharm* and *#selfinjury* hashtags. Similar existing studies used these tags in various investigations on platforms such as YouTube [8], [24] and Flickr [7]. Our data collection covered up to one month time period (from 02/02/2020 to 07/03/2020). The TAGS sheets retrieved 8,523 tweets with *#selfharm* hashtag producing up to 3,481 tweets, and *#selinjury* accounts for only 1,394 tweets. This produced a total of 4,875 unique tweets from 3,421 users. Around 94.5% of these tweets were posted using English language, with 2% representing Indonesian language while the remaining percentage were from other languages in Africa and some parts of Europe. However, our analysis focused on tweets written in the English language.

A. Data Analysis

In our preliminary analysis, we compared the frequencies of tweets produced by the two hashtags. The idea was to understand the most common and popularly used hashtag that members of this community use on Twitter. We observed that tweets containing *#selfharm* were more than double the number of tweets posted with *#selfinjury*. Meanwhile, Figure 2 shows the distribution of the tweets we retrieved using these

tags. Using a weekly interval, we discovered many tweets on 01/03/2020. This date is globally recognised as a self-injury awareness day. On that date, many of the users utilise Twitter by actively engaging in raising awareness about self-injury and other mental health-related information. On average, users posts 99 tweets per day using *selinjury* as opposed to *selfinjury* with only 39.

Moreover, we traced the number of tweets that people post per hour. Intuitively, tweeting about self-injury during the night time could be linked with lack of sleep and increases the likelihood of depression which is associated with self-injury. Figure 1 illustrates the summary of the frequently hours used in posting the examined tweets. We found that from 00:00 to 06:00 UK time, tweeters were less active in tweeting about NSSI information. From 08: 12 to 12:00 users became very active by sending many tweet messages discussing self-injury. There was a decrease in the number of tweets messages posted between 12:00 and 14:00, which indicates that there is a possibility that users could be on lunch break. Additionally, the number of tweets began to decline from 18:00 to 23:00. This indicates that most users were less active as they do not want to distract their sleep.

As suggested by [25], we performed text processing using Python by importing Natural Language Processing ToolKit (NLTK) library and other relevant libraries. The first step in processing our data was deleting any line character, removing numbers, punctuation, and hashtags. In doing so, we used regular expressions and also removed links and any non-alphabetical symbols. In the second step, we converted all tweet texts to lower cases and tokens and numbers. Similarly, we removed stop words as they are not informative in our text analysis and performed lemmatization to get the appropriate features. Moreover, we processed our data using two lists. In the first list, we have a set of words frequency while the second list contains a set of tokenized words for any piece of a document or tweet. Using these two lists, we constructed a Bag of Words (BoW) and created vectors to represent the documents. Intuitively, this results in a matrix table with rows as documents and columns containing the vocabularies. However, we do not consider performing stemming. This procedure could delete the last few characters of our tokens by producing incorrect spellings such as *peopl* for people.

Our data processing takes into account the size or length of tweets as provided by Twitter. This precaution will help to avoid in-depth text pre-processing as this could affect understanding meanings and context. As shown in Figure 3, only a few tweets are between 0 to 49, and 230 to 280 characters long. Twitter allows users to post a tweet that is not more than 280 character lengths [26], with this size, users can express their thoughts or opinion in brief form. The vast majority of the tweets we collected using the defined hashtags were between 150 to 200 characters.

B. Tweeter's Bio-Descriptions

Several existing studies analysed tweeters using content analysis that depends on qualitatively coding users account description. The work of [15] grouped Twitter users who discussed dementia using inductive coding schemes. Additionally, some authors used similar approaches and classified users who engaged in tweeting science articles on Twitter [27]. When representing users in our experiment, we only considered the " user bio" field of our dataset as Twitter uses this to represent the text users provided that described them. As we know, texts created by users on social networking sites often contained oddities. Again, because Twitter uses hashtags and @mention features, different unique characters appear in fields containing texts. Taking these into account, we first performed text processing on the user's bio descriptions to have clean data for our experiment.

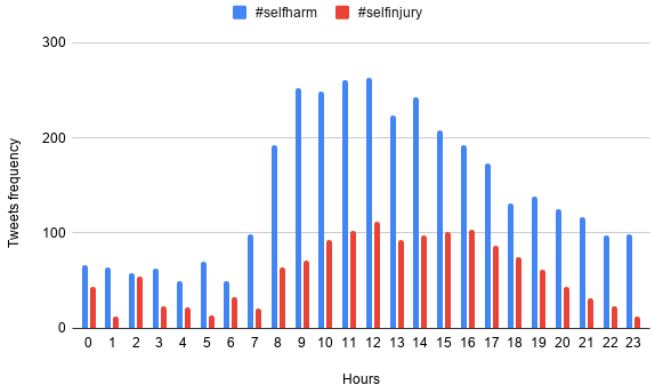

Fig. 1. Hourly tweets distribution.

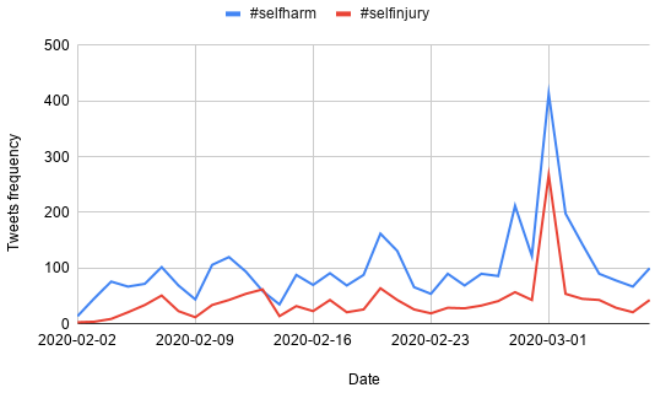

Fig. 2. Daily tweets frequency.

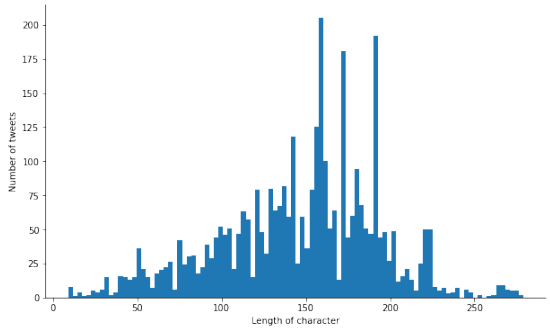

Fig. 3. Tweets character sizes.

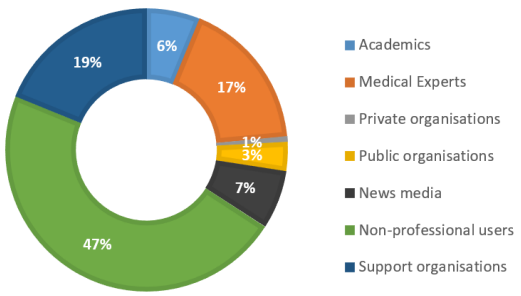

Fig. 4. Category of users .

Furthermore, we developed a codebook and inductively classified the tweeters into one of the groups mentioned in Table II. This content analysis approach is a suitable technique for categorizing Twitter users based on their profile descriptions [15]. One of the primary goals of using inductive coding in our analysis was to identify common groups that are immanent in our data, thereby maintaining the data [28]. Although there are different types of inductive coding, our study considered using summative analysis as this type focused on word frequencies, and their meanings in identifying underlying themes [28]. Nevertheless, our research implemented the inductive coding approach in grouping tweeters as suggested by [28]. In line with this, the first author and another researcher whose background is computer and information sciences worked on classifying tweeters into one of the listed sender group in Table II.

C. Opinion Mining

Meanwhile, one of the cutting edge techniques of examining users' opinions on social media text is through using Valence Aware Sentiment Reasoner (VADER) [29]. This rule-based approach measures the positivity or negativity sentiments of textual data posted on social media. After comparing VADER with current existing state-of-the-art techniques of analyzing text sentiments such as Linguistic Enquiry Word Count (LIWC), it was found that VADER surpassed other techniques in detecting people's opinion expressed in writings [29].

D. Topic Analysis

Latent Dirichlet Allocation (LDA) is among the current state-of-art techniques of topics analysis that viewed documents as a collection of multiple topics [19]. As a Bayesian model, the LDA consist of three hierarchical levels that each particular item in a set of the collection is viewed as a representation of definable mixture distributed over a set of topics [19]. In LDA algorithm, topics or themes of a text corpus are a set of words in a document and the distribution of these words provides room for observations and inferences [19]. The model offers a robust tool for identifying the unknown themes in substantial document collections. We applied LDA algorithms to examine themes surrounding the set of tweets discussing intentional self-injury on Twitter. We chose this algorithm because of its generic nature, and it was found to be effective in the existing studies. For example, some researchers proposed a novel approach for hashtag recommendation to facilitate micro-blog posts (tweets) search [30]. Their method depends on LDA to discover topics and tweets classifications.

IV. Results

A. Who is tweeting NSSI on Twitter?

Most of the studies that researchers have done on Twitter were focused on the tweets users created rather than actual profile descriptions of the users. Twitter allow users to disclose their professions and locations on their profile. We utilised this information to identify the categories of people discussing NSSI on Twitter using *#selfharm* and *#selfinjury* hashtags. However, our study approached this question through a mixed-method approach and explored different users' tweeting intentional self-injury. The first author and another researcher in the field computer and information sciences used a coding scheme (see appendix) and classified users into different categories. We used Cohen's kappa in measuring the reliability of inter-coder agreement [31]. We found a score of .72; this shows a strong level of agreement between them. The pie chart in Figure 4 illustrates the percentage of different classes of users discussing NSSI on Twitter.

We found that nearly half (47%) of the users are non-professionals. From the set of bigrams shown in Table II, many of these users are bloggers and health advocates. Team of medical experts have only 17% while the group of academic professionals is only 6% of the entire set of users. Although medical experts play an essential role in addressing self-injury behaviour, our analysis shows their proportion is not more than one-fifth of the examined users discussing self-injury on Twitter. Group of charity organisations that give online social supports to self-injurers reached up to 19%. This percentage is significantly higher than public and private organisations categories, with only 3% and 1%. Similarly, as seen from the pie chart on Figure 4, news agencies engaged in self-injury discussions on Twitter, and they represent 7% of the entire set of users.

B. What tweeters tweet about self-injury on Twitter?

Social networking sites like Twitter become an online space for individuals to disclose their experiences and opinions. It is very challenging to explore and discover useful insights on what people are tweeting concerning non-suicidal self-injury. Using unsupervised computational approaches, we can understand this sensitive information that users are discussing on Twitter. Our study applied the LDA algorithm and discovered various hidden topics in tweets containing self-harm tags. The difficulty of choosing the correct or appropriate number of clusters is one of the critical issues of applying LDA in topic modelling. This concern is addressed by computing the appropriate number of clusters using the following metrics;

$$Perplexity(D_{est}) = exp\left\{ -\frac{\sum_{d=1}^{M} logP(w_d)}{\sum_{d=1}^{M} N_d} \right\} \quad (1)$$

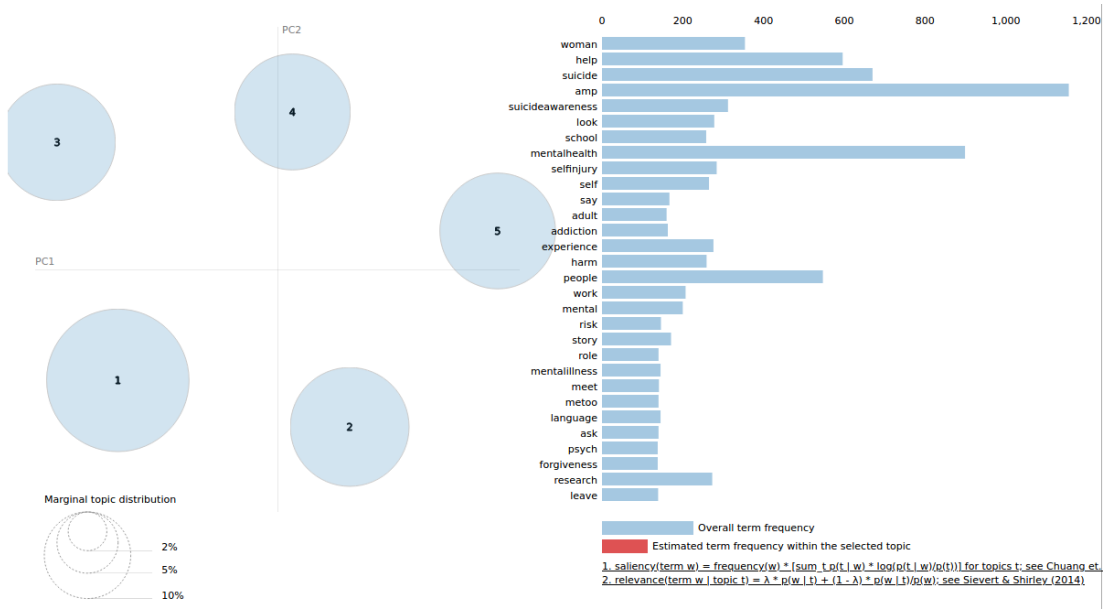

Fig. 5. LDAvis: Inter-topics distances and topics relevant terms.

$$L = log(P(wd|w_1, w_2, w_3..., w_n) = \left(\frac{logfreqw_1...}{wnfreq(w_1...w_{n-1})}\right) \quad (2)$$

From these equations, M denotes the total number of the documents, and P represents the probability of a word W in a document. Also, N stands for the number of words in each document. These metrics are very significant in experimenting with topics using probabilistic techniques. Usually, the pattern of topics in text documents is invisible. Consequently, we can only analyze the documents and words to gain useful information.

Our experiment ran different LDA models on several topics (5, 10, 15, 20, 30) and compared the perplexity and likelihood scores across the models. Consequently, the model with five topics produced the lowest perplexity value, which is **455.60** and a high likelihood of **-160034.81**. Intuitively, when the perplexity value is low, and the likelihood is high, the model is considered effective. Moreover, we validate and visualise this model using LDAvis, the contemporary technique for interpreting and displaying topic models [32]. LDAvis is an interactive web-based system that allows for more in-depth word inspections and topics relationships. The tool enables researchers to view topics similarities as well as differences in a more interactive manner. From the left-hand side of Figure 5, we have our model's broad topics presented in circles. For each circled topic, its area is proportional to the total number of topics in the corpus.

Figure 8 shows the inter-topics distances computed using Principal Component Analysis (PCA). The horizontal bar chart on the right-hand side of the figure presented the estimated term frequencies in each topic and the entire corpus's overall term frequency. For any chosen topic, its relevant terms are displayed on the red colour-coded bar graph. For instance,

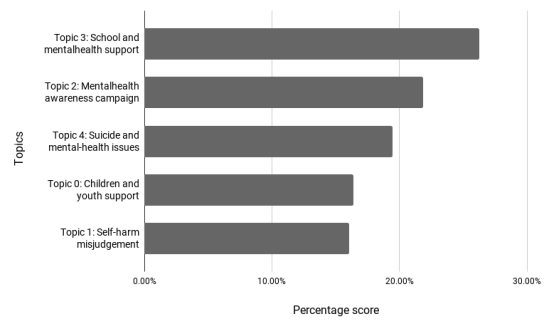

Fig. 6. Percentage of topics.

when topic five was selected, its relevant terms are sorted and ranked from higher to lower probabilities of relevance in the red bar.

Using the visualisation of the LDAvis technique, we selected the top 10 most relevant terms from each topic and interpreted the topics as; *children and youth well-being*, *self-harm misjudgement*, *mental health awareness*, *school and mental health support* and, *suicide and mental-health issues*. Although LDA assigns these topics to each document, only one topic could occupy a large portion of a document. Our experiment calculated the total number of topics from the tweets collection and found the most dominant topic that users are discussing on Twitter. Figure 6 shows that around 26.26% of the conversations were directed towards helping young people who are self-harming and the need to provide mental health support in schools. Subsequently, up to 21.83% of these discussions suggest offering mental health awareness with a focus on women group and recommend other critical sources of online support such as the Harmless UK.

Despite many arguments about the connection between self-harm and suicide, approximately 19.45% of the investigated

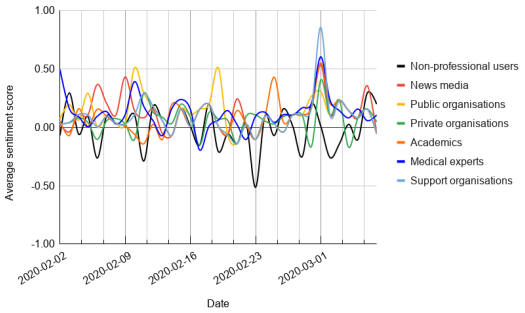

Fig. 7. Users group sentiment analysis.

tweets discussed supporting people with self-injury and suicidal thoughts. Meanwhile, topic one and two were the less dominant topics as the duo represent only 16% of the overall discussions. Table I illustrates the example tweet for each topic, and Figure 9 presents the distributions of these topics in the UK.

1) Mapping topics with locations: Even though some Twitter users prefer to remain anonymous, Twitter allows people to activate or deactivate a geo-coordinate feature that automatically records the city or country they are when using the platform. Users prefer to write their location in the bio-description rather than activating location coordinates feature. We analysed users' reported location information and obtained the list of countries to understand the topics users discuss from different places.

Our investigation uncovered that these topics are widely spread in the United Kingdom and some parts of Europe. For this reason, we focused on tweets posted in the UK. The idea of investigation UK regions is to understand the areas or places that are less engaging on Twitter in raising awareness about self-harm behaviour. Our study discovered that the percentage of all topics in England is significantly higher than the other counties. In terms of disseminating information about self-injury, this location is socially active on Twitter as all the topics account for more than 50%.

Furthermore, topic two and five were not much discussed in contrast to other topics. On the other hand, Scotland is the second place in the UK engaged in discussing self-injury on Twitter. Topic number two is the most prevalent topic discussed in Scotland with up to 18%. Consequently, we observed equal rates among all topics discussed in Wales and this is closely similar to Ireland and Northern Ireland.

C. What is the opinion of users tweeting self-injury on Twitter?

Twitter users express their opinions and emotions in tweets that are not more than 280 characters. Following our analysis on users' profile and categorisation, we further expanded our study by examining how these groups of users convey their feelings on Twitter. Their personal opinions regarding NSSI could be positive or negative, which could reinforce or reduce self-injury behaviour. Using VADER, we calculated each tweet's sentiment score and grouped the tweets according to the category of users. We computed each group of tweets' average sentiment value per day and plotted the result on Figure 7.

TABLE I
TOPICS WITH EXAMPLE TWEETS

Topic	Example tweet
Topic 0	*"Just had a great planning meeting for our #BodyMapping workshops! We are using body mapping to help youth visually represent their worlds. We will be using it in our #Research projects with youth who #SelfHarm; and with Indigenous youth aging out of the #ChildWelfare system."*
Topic 1	*"Sitting in a coffee shop. Remove jacket because it's warm & overhear a kid ask his mum what my #selfharm scars are. Her reply 'don't look at them, she's a freak'."*
Topic 2	*"This #selfharm awareness month. Learn the signs so you can help a friend or loved one self-harming. Be supportive, don't judge, learn more and help them get help."*
Topic 3	*"Today is #SelfHarmAwarenessDay & we are sharing a page from our ' Supporting Mental Health & Wellbeing in Schools' booklet to help school staff to support young people who #selfharm."*
Topic 4	*"I have a mental illness. I will die with it but, I refuse to die FROM it !!"*

From the chart, we discovered that negative opinion comes mostly from non-professional users and public organisations. Although negative sentiments have sometimes been received from medical experts, news media handles and support organisations, our results suggest that these groups have been dominant in issuing positive tweets related to self-injury. Finally, our experiment shows that public organizations sound more positively than private organizations regarding NSSI discussions on Twitter.

V. CONCLUSION

In this research, we used *"#selfharm"* and *"#selfinjury"* hashtags and retrieved tweets from Twitter social networks. Our research found some groups of users who participated in the NSSI conversations on Twitter. Even though non-professional users dominated nearly half of the users involved in the discussions, our analysis uncovered that medical teams are reaching self-injurers on Twitter to provide social support and suggest ways of recovery. We discovered that academic professionals are using Twitter to raise awareness of the dangers of NSSI and encourage self-harming individuals to participate in their research investigations. The sentiments expressed between different categories of users varies significantly. Support organisations, medical experts and academic professionals opinionated more positively. There is a large proportion of negative sentiments from Non-professional users.

Our study also investigated the nature of NSSI discussions by using the current state-of-the-art LDA topic modelling algorithm. Although users discussed how people misjudged individuals that are self-harming, the common dominant topics were directed on providing mental health support in schools and campaign awareness about NSSI. This demonstrates the positive use of Twitter social network in fighting self-injury

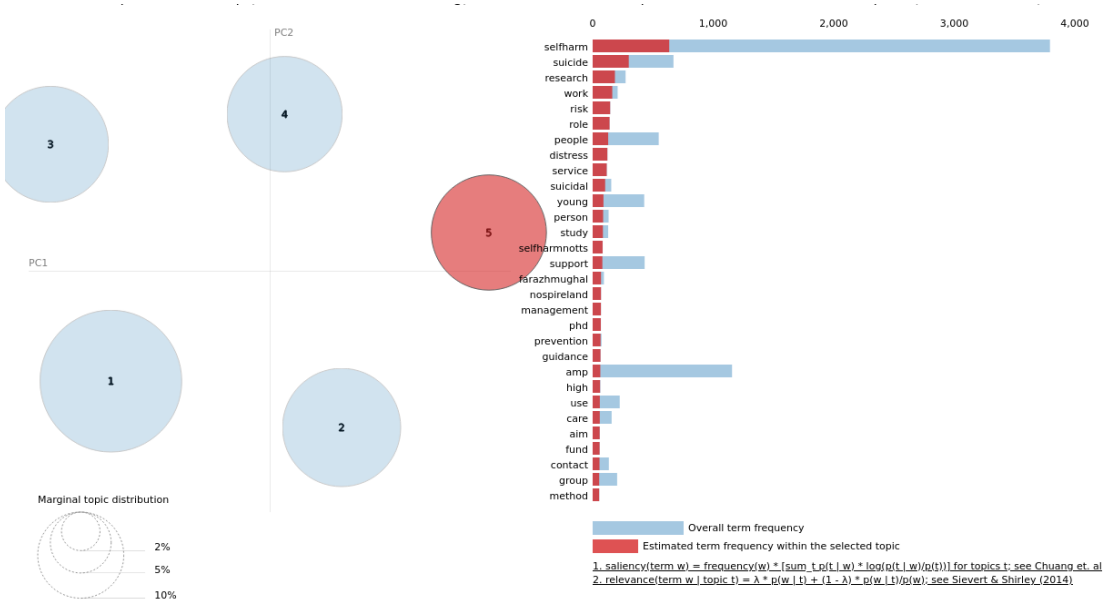

Fig. 8. Topic 5 and its most relevant terms.

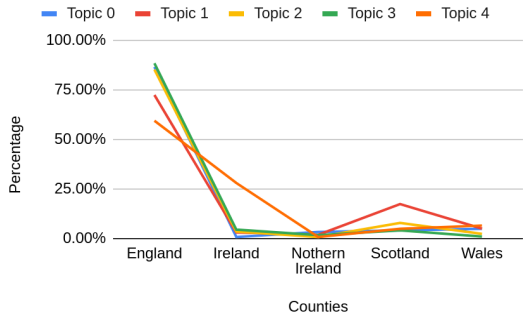

Fig. 9. Percentage of topics in the UK regions.

behaviour. Given that the number of negative sentiments from the non-professional users is significantly higher, future research will examine this group to determine the underlying causes. There could be a tendency that some members of this group have encountered triggering contents that may put them at risk of self-harm.

References

[1] A. Mahmood, "Psychology and shia muslims," *The psychologies in religion: Working with the religious client*, vol. 237, 2006.

[2] M. A. Moreno, A. Ton, E. Selkie, and Y. Evans, "Secret society 123: Understanding the language of self-harm on instagram," *Journal of Adolescent Health*, vol. 58, no. 1, pp. 78–84, 2016.

[3] E. M. Miguel, T. Chou, A. Golik, D. Cornacchio, A. L. Sanchez, M. DeSerisy, and J. S. Comer, "Examining the scope and patterns of deliberate self-injurious cutting content in popular social media," *Depression and anxiety*, vol. 34, no. 9, pp. 786–793, 2017.

[4] R. C. Brown, T. Fischer, A. D. Goldwich, F. Keller, R. Young, and P. L. Plener, "# cutting: Non-suicidal self-injury (nssi) on instagram," *Psychological medicine*, vol. 48, no. 2, pp. 337–346, 2018.

[5] N. Shanahan, C. Brennan, and A. House, "Self-harm and social media: thematic analysis of images posted on three social media sites," *BMJ open*, vol. 9, no. 2, p. e027006, 2019.

[6] F. Arendt, S. Scherr, and D. Romer, "Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults," *New Media & Society*, vol. 21, no. 11-12, pp. 2422–2442, 2019.

[7] Y. Wang, J. Tang, J. Li, B. Li, Y. Wan, C. Mellina, N. O'Hare, and Y. Chang, "Understanding and discovering deliberate self-harm content in social media," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 93–102.

[8] S. P. Lewis, N. L. Heath, M. J. Sornberger, and A. E. Arbuthnott, "Helpful or harmful? an examination of viewers' responses to nonsuicidal self-injury videos on youtube," *Journal of Adolescent Health*, vol. 51, no. 4, pp. 380–385, 2012.

[9] S. P. Lewis and A. K. Knoll, "Do it yourself: Examination of self-injury first aid tips on youtube," *Cyberpsychology, Behavior, and Social Networking*, vol. 18, no. 5, pp. 301–304, 2015.

[10] A. Khasawneh, K. C. Madathil, E. Dixon, P. Wiśniewski, H. Zinzow, and R. Roth, "Examining the self-harm and suicide contagion effects of the blue whale challenge on youtube and twitter: Qualitative study," *JMIR Mental Health*, vol. 7, no. 6, p. e15973, 2020.

[11] C. Emma Hilton, "Unveiling self-harm behaviour: what can social media site twitter tell us about self-harm? a qualitative exploration," *Journal of clinical nursing*, vol. 26, no. 11-12, pp. 1690–1704, 2017.

[12] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist, "Understanding the demographics of twitter users." *Icwsm*, vol. 11, no. 5th, p. 25, 2011.

[13] L. Sloan, J. Morgan, P. Burnap, and M. Williams, "Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data," *PloS one*, vol. 10, no. 3, p. e0115545, 2015.

[14] D. Preoţiuc-Pietro, V. Lampos, and N. Aletras, "An analysis of the user occupational class through twitter content," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1754–1764.

[15] F. Alhayan and D. Pennington, "Twitter as health information source: exploring the parameters affecting dementia-related tweets," in *International Conference on Social Media and Society*, 2020, pp. 277–290.

[16] A. D. Damiano and J. R. Allen Catellier, "A content analysis of coronavirus tweets in the united states just prior to the pandemic declaration," *Cyberpsychology, Behavior, and Social Networking*, 2020.

[17] S. E. Victor and E. D. Klonsky, "Daily emotion in non-suicidal self-injury," *Journal of Clinical Psychology*, vol. 70, no. 4, pp. 364–375, 2014.

[18] S. Rasmussen, K. Hawton, S. Philpott-Morgan, and R. C. O'Connor, "Why do adolescents self-harm?" *Crisis*, 2016.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[20] B. Dahal, S. A. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Social Network Analysis and Mining*, vol. 9, no. 1, p. 24, 2019.

[21] A. Karami, L. S. Bennett, and X. He, "Mining public opinion about economic issues: Twitter and the us presidential election," *International Journal of Strategic Decision Sciences (IJSDS)*, vol. 9, no. 1, pp. 18–28, 2018.

[22] M. J. Paul and M. Dredze, "You are what you tweet: Analyzing twitter for public health," in *Fifth international AAAI conference on weblogs and social media*. Citeseer, 2011.

[23] M. Hawksey, "Twitter archiving google spreadsheet tags v5," *Retrieved from The Musing of Martin Hawksey (EdTech Explorer): https://mashe. hawksey. info/2013/02/twitter-archive-tagsv5*, 2013.

[24] M. A. Alhassan and D. Pennington, "Detecting critical responses from deliberate self-harm videos on youtube," in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 2020, pp. 383–386.

[25] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

[26] K. Gligorić, A. Anderson, and R. West, "How constraints affect content: The case of twitter's switch from 140 to 280 characters," *arXiv preprint arXiv:1804.02318*, 2018.

[27] J. Vainio and K. Holmberg, "Highly tweeted science articles: who tweets them? an analysis of twitter user profile descriptions," *Scientometrics*, vol. 112, no. 1, pp. 345–366, 2017.

[28] D. R. Thomas, "A general inductive approach for analyzing qualitative evaluation data," *American journal of evaluation*, vol. 27, no. 2, pp. 237–246, 2006.

[29] C. Gilbert and E. Hutto, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*, vol. 81, 2014, p. 82.

[30] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen, and R. Van de Walle, "Using topic models for twitter hashtag recommendation," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 593–596.

[31] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[32] C. Sievert and K. Shirley, "Ldavis: A method for visualizing and interpreting topics," in *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 2014, pp. 63–70.

APPENDIX A
USER CATEGORISATION CODE

TABLE II
ANNOTATION GUIDE AND EXAMPLE BIGRAMS

Tweeter User Group	Descriptions	Top bigrams
Academics	There are Twitter handles of academic research professionals who are from universities and other tertiary institutions. This category is mainly for users that reported to be experts in any field of study from academic institutions. Examples are professors, lecturers, and researchers.	mental health, phd student, phd researcher, research fellow, selfharm suicide, suicide prevention, young people, health psychology, psychologist research, research selfharm.
Medical experts	Individual users who post information about NSSI and have written in their bio-description that they are health care workers. Examples are medical doctor, nurse, and psychiatrist.	mental health, clinical psychologist, suicide prevention, psychiatry epidemiology, neuro psychiatry, psychiatry epidemiology, health nurse, educational psychologist, nurse independent, personal capacity.
Non-professional users	In this group, we considered a Twitter account handle belonging to one particular person instead of an organisation or a joint account participating in NSSI discussions. This group is for users that are not medical or academic professionals. Examples are those who claimed to be artist, blogger, celebrity, or parents of self-injurers.	mental health, public speaker, passionate independent, health advocate, public health, advocate blogger, self support, mental illness, social justice, young people.
Public Organizations	This category represents a description of a Twitter handle which contained information that shows government organizations account. In this group, we listed twitter accounts that have clearly described themselves as public agencies such as public schools, public medical institutions, and research centres (run by the government).	school workshop, world school, school international, international school, provide learn, learn great, information resource, wellbeing information, resource free, young people.
Private Organizations	This cohort is for any Twitter handle that declared on its profile that, the account belongs to a private organization that is advertising or selling products to the general public for business purposes. Example of these products could be mental health or self-harm applications designed to support people for effective well-being and self-harm recovery.	mental health, help create, create happy, happy productive, productive working,working environment, training healthcare, healthcare socialcare, expert workplace, centre excellence.
Support organisatioons	Several charity organisations educate users on the nature of self-harm and other mental health problems. Such organisations receive support from the community or donations. This group is for account handles that provide information and advice related to mental health and NSSI. Examples include Samaritans, young minds, and mind charity.	peer support, mental health, support mentalhealth, suicide selfharm, selfharm awareness, awareness response, response safetyplan, safetyplan just, information resource, mentalhealth depression.
News Media	This group is mainly for Tweet sender bio descriptions explicitly reported being owned and managed by any recognised news media companies. In other words, this category is for both national and international media Twitter accounts. Examples could be the British broadcasting corporation (BBC), the guardian, and the telegraph news media.	substance abuse, purpose help, help prevent, voice population, update news, medium podcast, podcast live, live online, online mentalhealth, mentalhealth event.
Others	This group is for any undecipherable account whose description is provided in a language other than English. A Twitter handle that failed to meet any of the categories' criteria described above is considered in this cohort.	