Future Generation Computer Systems 93 (2019) 971–978

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in Twitter

Raúl Lara-Cabrera *, Antonio Gonzalez-Pardo, David Camacho

Computer Science Department, Universidad Autónoma de Madrid, Spain

HIGHLIGHTS

- Risk assessment indicators and metrics are analysed on this paper.
- Keyword-based metrics perform well when it comes to highlight radicalised users.
- It is possible to assess the risk of radicalisation analysing a user's tweets.

ARTICLE INFO

Article history:
Received 30 April 2017
Received in revised form 20 July 2017
Accepted 26 October 2017
Available online 4 November 2017

Keywords:
Social Network Analysis
Risk assessment
Complex networks
Radicalisation factors

ABSTRACT

Nowadays, Social Networks have become an essential communication tools producing a large amount of information about their users and their interactions, which can be analysed with Data Mining methods. In the last years, Social Networks are being used to radicalise people. In this paper, we study the performance of a set of indicators and their respective metrics, devoted to assess the risk of radicalisation of a precise individual on three different datasets. Keyword-based metrics, even though depending on the written language, performs well when measuring frustration, perception of discrimination as well as declaration of negative and positive ideas about Western society and Jihadism, respectively. However, metrics based on frequent habits such as writing ellipses are not well enough to characterise a user in risk of radicalisation. The paper presents a detailed description of both, the set of indicators used to assess the radicalisation in Social Networks and the set of datasets used to evaluate them. Finally, an experimental study over these datasets are carried out to evaluate the performance of the metrics considered.

1. Introduction

As a social structure that consists of actors and the interrelationships between them, Social Networks are inherently related to the human being, although their study did not become popular until the 1930s. Nowadays, Social Networks have escalated and increased their size due to the penetration of the Internet in the society, hence simplifying their analysis due to the high opening of data as well as their high volume.

Social Network Analysis (SNA) emerged as a set of methods and tools for the analysis of social structures and the investigation of the social aspects of these structures [1]. As an interdisciplinary research field, SNA together with other complex networks became a major paradigm in sociology as well as in other formal sciences [2].

An interesting application of SNA is the detection and extraction of behavioural patterns from data obtained from Social Networks. By combining metrics at different levels of the network (macroscopic, microscopic and mesoscopic) as well as by studying the evolution of relationships between individuals, behavioural models can be built that, somehow, represent the members of the network.

In the last years, the world in general and the West in particular are in a situation of constant danger due to the terrorist attacks perpetrated by the so-called Islamic State of Iraq and Syria (ISIS). It is not surprising that the fight against terrorism has become one of the priority objectives of any country.

Jihadist terrorism has a distinct feature compared to other kind of terrorism: the way they radicalise and recruit their militants. People usually begin to be radicalised by contacting radical individuals or groups, which provides feel of belonging and social recognition. Moreover, radicals tend to publish a lot of information on the Internet, specifically in Social Networks such as Twitter, Tumblr, Facebook, Instagram and Youtube [3,4].

Therefore, it is interesting to apply techniques of SNA to the data published by jihadist radicalised people to extract behavioural patterns that, in turn, help to assess the risk of radicalisation.

The goal of this paper is to study on a set of indicators devoted to risk radicalisation assessment and their respective metrics used

* Corresponding author.
E-mail addresses: raul.lara@uam.es (R. Lara-Cabrera), antonio.gonzalez@uam.es (A. Gonzalez-Pardo), david.camacho@uam.es (D. Camacho).

https://doi.org/10.1016/j.future.2017.10.046
0167-739X/

to quantify them. These indicators come from several experts psychologists on radicalisation and are focused on the behaviour expressed in Social Networks. Furthermore, these indicators are easy to compute in an automated way, gathering and analysing messages from Social Networks. This should improve the present radicalisation risk assessment process, which is frequently done using manual tools such as VERA [5,6].

The main contributions of the present paper are the following:

- Five indicators to assess the risk of radicalisation as well as the metrics to quantify them are presented and studied.
- Three datasets have been built in order to analyse the performance of the aforementioned metrics.
- The results obtained provide guidance on how to improve the present indicators in order to achieve better performance when assessing risk of radicalisation in Social Networks.

The remainder of the paper is structured as follows. First a brief introduction of the principles of complex networks and what are the most commonly used algorithm to analyse these networks can be found in Section 2. Then, the different indicators and the datasets used in this work are described in Section 3. Section 4 provides a detail description of the experimental phase that evaluate the performance of the metrics. Finally, the conclusions drawn from the experimental phase and future research works can be found in Section 5.

2. Background

Complex networks have been studied mainly in the form of mathematical graph theory and also in the social sciences [7]. Nowadays, the focus have shifted to networks with a high number of vertices and edges, trying to analyse their large-scale statistical properties. The presentation of models for scale-free [8] and power-law [9], as well as the study of new algorithms related to complex networks such as clustering [10,11] and community finding [12–14], have led to the study of many different issues in this sense.

There are many works related to the study of different types of complex networks in the literature. For instance, the information network made of citations between academic papers in which every node is a paper and there is a directed edge if a paper cites another [15]. Other approach is the one followed by [16–18], in these works the authors of the academic papers are represented in the nodes of the co-authorship network and the different edges imply there is at least an academic paper written by them. Regarding Social Networks [19], there are studies on business relationships between companies [20], supply chain context [21] and obesity [22], to name a few.

Concerning terrorism and Social Networks, there are several approaches in the literature to tackle this problem. For instance, Wadhwa and Bhatia [23] proposed a Data Mining approach for detecting the dynamic behaviour of radicals by analysing the messages posted. The approach consists of a message filtering and preprocessing stage followed by a topic identification algorithm, and other SNA techniques such as community detection and identification of key nodes.

O'Callaghan et al. [24] performed a detailed analysis of the communities found in messages and videos from Twitter and Youtube regarding the Syria conflict. Their findings indicate that social media activity in Syria was more complicated than that reported by other studies. Furthermore, the authors studied the effect that certain key events had over the number of videos uploaded to Youtube.

Expanding the focus, Ferrara et al. [25] designed a Machine Learning approach to detect extremist users levering temporal, network and meta-data features. In addition, this approach is able to predict content adopters and interaction reciprocity in social media. A similar approach was used by Agarwal and Sureka [26], who used a Support Vector Machine (SVM) classifier to detect online radicalisation on Twitter, following a semi-supervised learning scheme. Furthermore, Ashcroft et al. [27] followed a Machine Learning approach as well, in order to classify a tweet as containing material that is supporting jihadists groups or not. For its part, Kaati et al. [28] used the AdaBoost classifier to detect tweets that disseminates jihadist propaganda through the social media.

On the other hand, Glowacki et al. [29] show that the formation of raids for intergroup violence depends on the presence of specific leaders who tend to occupy a central position in the social network driven by the friendship relationship.

For a more comprehensive review of the literature, Correa and Sureka [30] published a survey on solutions to detect and analyse online radicalisation.

3. Materials and methods

This section is devoted to describe a set of indicators to assess the risk of radicalisation as well as to explain the datasets we have used to study the performance of the aforementioned indicators.

3.1. Indicators

As previously mentioned, the goal of this paper is to study the efficiency of different indicators that highlight those Social Network users with high risk of being radicalised. The studied indicators come from several expert psychologists on radicalisation and terrorism that are used to assess the radicalisation risk of an individual. Although it is possible to use many indicators on this context (see [5,6,31]), we focused on a manageable set of indicators which, in turn, are easily measurable using Social Networks data.

The set consists of **five indicators** grouped in two categories: *personality* and *interpersonal* relationships, and *attitudes* and *beliefs* towards Muslim religion and Western society. The former contains those indicators related to the writing style specific for each user, whereas the later are indicators are measured by the content of the tweets:

- **Personality related Indicators**:
  - **I1** *The individual is frustrated*. To measure this indicator in Social Networks we will take into account some aspects such as swearing, writing sentences fully capitalised and using words with negative content.
  - **I2** *The individual is introverted*. Aspects to be analysed: using ellipses (i.e. …) in the messages and measuring their length (introverted people tend to write short sentences).
- **Attitudes and beliefs related Indicators**:
  - **I3** *Perception of discrimination for being Muslim*. This perception may be expressed in the messages using some keywords related to discrimination.
  - **I4** *Expressing negative ideas about Western society*. As occurs with discrimination, it will be used several keywords related to negative ideas when anybody talks about the Western life style.
  - **I5** *Expressing positive ideas about jihadism*. Radical people show support and positive ideas about those engaged in the Jihad. Again, it is possible to analyse the usage of keywords related to this issue.

Please refer to [32] for a deeper explanation on the indicators and their operativeness (how they can be measured in a quantitative way), and Table 1 for the initial set of keywords used for each indicator.

Table 1
Initial keywords used to evaluate each indicator, they have been expanded to a larger set of keywords with synonyms obtained from Wordnet, and by seeking the occurrences of the stem of the words (stemming process). This is a two-step process: first keywords are expanded with Wordnet and then we keep their stems, which is what the method look for in the tweets.

Indicator	Initial keywords
I1. The individual is frustrated.	Shit, crap, damn, fuck.
I1 Use of words with negative content.	Hate, guilt, shame, terrible, horrible, bad, fault.
I3. Perception of discrimination for being Muslim.	Muslim, sick,hate, discrimination, people, racism, religion.
I4. Expressing negative ideas about Western society.	Western, hate, suck, people, west, europe, usa, US, bloody, sick, impure, kuffar, kafir.
I5. Expressing positive ideas about jihadism.	Islamic,state, caliphate, rise, mujahideen, mujahid, help, fight, weapon, gun, weapons.

Table 2
Top 10 writing languages detected in Anonymous dataset.

	Language	Percentage (%)
1	Arabic	**88.98**
2	Russian	4.45
3	Turkish	2.23
4	US English	*1.55*
5	French	1.30
6	Dutch	0.91
7	UK English	*0.10*
8	Urdu	0.08
9	Bulgarian	0.06
10	Hindi	0.05

3.2. Datasets

D1. Due to the special sensitiveness of this paper's topic, it is really difficult to find a publicly accessible dataset[1] . Authors in the literature usually build their datasets on their own and they rarely publish them. In a previous work [32] we used a dataset that, as far as we know, is the only open published dataset including Twitter usernames and tweets from ISIS sympathisers. Furthermore, this dataset was built and curated by a digital agency that serves government agencies and is available at Kaggle[2] with the following description: "We scraped over 17,000 tweets from 100+ pro-ISIS fan-boys from all over the world since the November 2015 Paris Attacks". Hence, this dataset should be considered as the most reliable.

D2. Other dataset was the one built by volunteers all around the world from the so-called Anonymous collective that gathered and published many Twitter accounts of people related to ISIS during the `#OpISIS` operation [33]. Several dump files can be found at Pastebin[3] , a web platform frequently used to publish this kind of information. Unlike with the previous one, this dataset has not been validated by experts but it was the same anonymous users who reported the accounts in Twitter with a certain hashtag and somebody grouped them in a dump file. So the reliability of this dataset should be taken with caution. One of the relevant characteristics of this dataset is related to the number of languages used in the tweets, several languages as English, Arabic, Russian, Turkish, French, have been employed to write them. Most of the tweets have been written in Arabic language, whereas English language are used only in a minor proportion (1.65% between both US and UK English). Next Table 2, shows a distribution of these tweets by language.

Once the dump files were downloaded, a tool developed in Python downloaded information about users and their tweets by making requests to the Twitter API. The design of this tool took into account the limitations of the Application Programming Interface (API), such as the maximum number of requests that is limited in time as well as the number of tweets that can be downloaded from a user in each request.

Table 3
Summary of the datasets used in this work.

	D1	D2	D3
Number of users	112	142	120
Number of tweets	17 410	76 286	173 530
Avg. tweets per user	155.45	527.23	1446.08
Stdev. tweets per user	269.54	901.48	1114.92

D3. To complement the above and analyse the performance of the metrics computed from the indicators we built another dataset. This new dataset represents a set of tweets from users that were randomly selected (so it can be used to evaluate any algorithm or metric, comparing random data against the data target). More precisely, we attached a software crawler to the streaming API of Twitter during 30 s and then we randomly selected 120 users. To avoid any bias on the dataset, we did not use any keyword nor hashtag in the query, because we were trying to obtain a dataset as randomised as possible. Finally, we gathered tweets published by the aforementioned users.

Table 3 shows a summary of previous datasets, the number of users and total amount of tweets are shown. The number of users are quite equivalent for all the datasets considered (around a hundred), the number of tweets have been increased for the random dataset (D3). Finally, to better understood the structure of these networks, the average number of tweets and the standard deviation for each user have been included. As it could be expected, the random dataset has the higher average of tweets per user. This is a side effect of the limitations imposed by the Twitter's API: it is only possible to obtain, at most, up to 3200 of a user's most recent Tweets, so it makes sense that the average number of tweets for a random sample is around the half point within that range.

4. Analysis of the metrics over the datasets

This section studies and analyses the performance of the different indicators over the datasets. These indicators have been operationalised (measured) using several metrics, in form of specific kind of words as swear/negative ideas or keywords, which can be extracted from the tweets to be later statistically processed.

4.1. Experimental setup

To study the performance of the indicators, we followed a process that involved several stages (see Fig. 1). The first one, data preprocessing, homogenises all the datasets as well as the messages from every user, removing the URLs and mentions using regular expressions. This is performed due to the low information this items provide to the indicators as they are defined.

[1] All of the datasets used in this work will be publicly available. They can be downloaded from: http://aida.ii.uam.es/resources.

[2] https://www.kaggle.com/kzaman/how-isis-uses-twitter (Last accessed: April, 2017).

[3] https://pastebin.com/u/CyberRog (Last accessed: April, 2017).

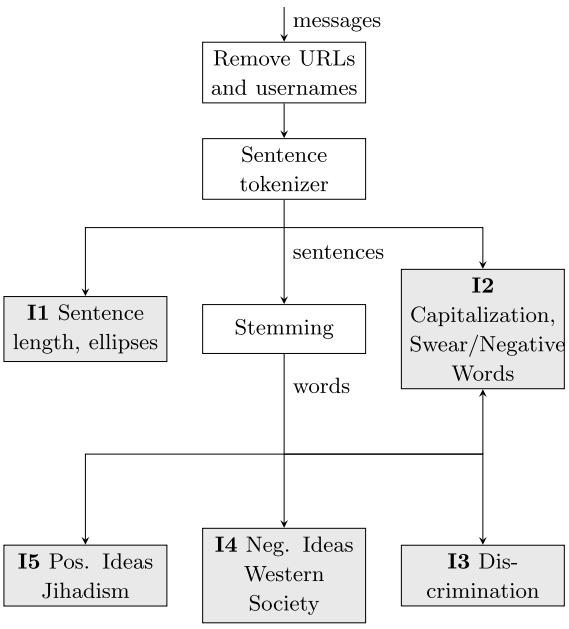

Fig. 1. Data flow: preprocessing, cleaning and analysis to compute indicators.

As indicators **I1** and **I2** are based on sentences, messages are then tokenized into this unit of language by taking into account the punctuation marks and new lines found in the messages.

The remaining indicators (**I3**, **I4** and **I5**) are computed by counting keywords related to the indicator, we decided to widen the search by expanding the set of keywords with their respective synonyms querying WordNet [34], and by seeking the occurrences of the stem of the words (stemming stage). The latter process avoids unwanted situations as not counting as an occurrence the word and its plural form, and is generally known as stemming.

Once the metrics associated to the indicators were computed for the three datasets, it have been analysed the distribution of these values and also compared them among the datasets. Note that metrics are quantitative measures related to some indicator in a many-to-one relationship, thus an indicator may comprise several metrics. Following, the results for each indicator are described and analysed individually.

4.2. Frustration

To measure the frustration of any user we focused on two metrics: *swearing* and the usage of *words with negative content*. They are computed by counting the frequency of their respective keywords, that is, the number of times a keyword appears in a tweet.

As shown in Fig. 2(a), most of the users in D2 have a low swearing ratio, that is, the number of times a tweet contains a keyword divided by the user's total number of tweets. This may be due to the fact that most of the Twitter accounts reported by Anonymous did not use English as the written language.

On the other hand, users from D1 and D3 exhibit a similar behaviour regarding the usage of swear words, with a rather similar number of users having a high swearing ratio. This suggests the inability of swear words to characterise individuals in risk of radicalisation. However, median swearing ratios in datasets D1 and D3 were 0.039126 and 0.001010 respectively, so there is a significant evidence that users in risk of radicalisation tend to have a higher swearing ratio than random users (*Wilcoxon rank sum test*: $W = 8261.5, n_1 = 112, n_3 = 120, p\text{-value} = 0.001025$, one-tailed).

Regarding to the use of words with negative content, it can be observed a similar effect as with swearing (see Fig. 2(b)): it shows a low ratio on accounts of D2, and a slightly higher density ratio on users from D1 with respect to those in D3. Again, the median negative content ratios of 0.027507 and 0.0 in D1 and D3, respectively, shows a statistically significant evidence that former users tend to use more words with negative connotations than the latter (*Wilcoxon rank sum test*: $W = 9142.5, n_1 = 112, n_3 = 120$, p-value $= 3.715e-07$, one-tailed).

These results suggest that analysing the use of words with negative connotations as well as swearing, are good metrics to measure the frustration of an individual regarding his/her risk of being radicalised. In other words, it seems that anyone in risk of radicalisation is more likely to use negative and swearing words than the average twitter user. However, this should not be taken as a binary decision, but it could be used jointly with other risk radicalisation components to generate a more accurate prediction.

4.3. Introversion

To measure the introversion of the user we defined two metrics: the ratio of tweets that contain an *ellipsis* and the *median length* of

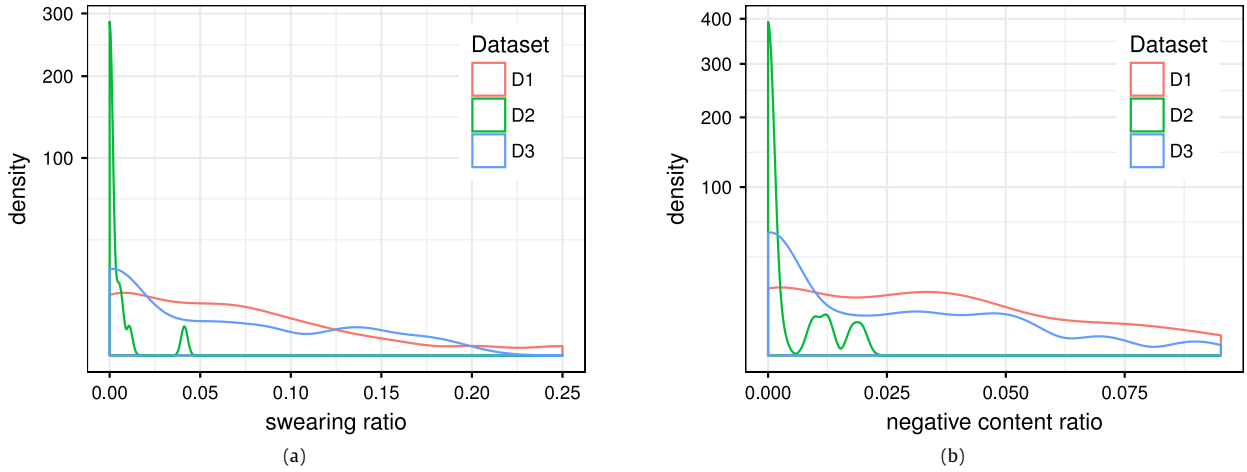

Fig. 2. Density distribution of the swearing (a) (exponential y-scale) and negative content (b) ratios for each user and dataset.

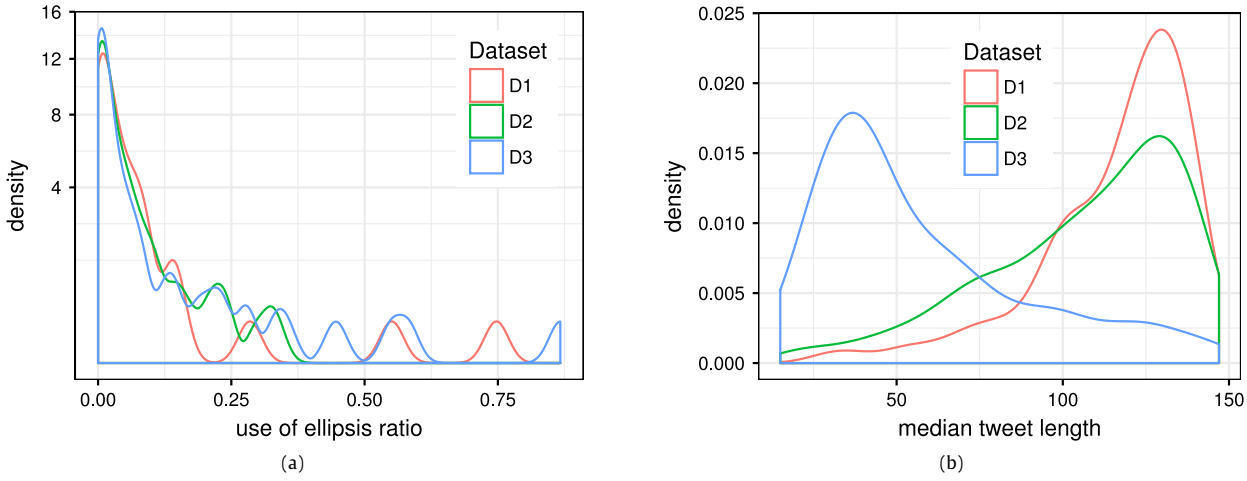

Fig. 3. Density distribution of the ellipses ratio (a) and median tweet length (b) for each user and dataset (exponential scale on both y-axis).

the user's tweets, computed as the number of characters and hence limited to 140 characters.

Fig. 3(a) shows the distribution of the ellipsis ratio for every user and dataset. It seems that the three distributions have the same behaviour and then the ratio should not be used to measure the indicator. In fact, we were unable to find significant evidence that the ellipsis ratio was be defined by different probability distributions (*Wilcoxon two-tailed rank sum test*: $W_{12} = 7933$, $W_{13} = 6777.5$, $W_{23} = 8600.5$, $n_1 = 112$, $n_2 = 142$, $n_3 = 120$, p-values $= \{0.9744, 0.9105, 0.8952\}$). This result makes sense as it is really common to use ellipsis when writing messages in a Social Network.

We found an interesting result regarding the median tweet length per user (see Fig. 3(b)): contrary to what the indicator suggests, users of the random dataset D2 write tweets shorter than those at risk of radicalisation (dataset D1 and D2), with medians 43.5, 123, 114 respectively. Furthermore, this difference in the median length of the tweets is statistically significant according to one-tailed Wilcoxon rank sum tests with $W_{31} = 1355$, $W_{32} = 2526.5$, p-values < 0.05.

These results suggest refocusing those metrics associated with the introversion indicator **I2**, since they are not able to establish a difference between users at risk of radicalisation and random users (use of ellipsis) or to measure the expected value (length of tweets). This makes sense: due to the limited length of tweets (140 characters), users tend to write many ellipses as a short way of enumerating things instead of writing them. Regarding the length of the tweet, as opposite to the expected, users in risk of radicalisation write longer tweets than the average user, pointing out the necessity of refocusing this metric.

4.4. Perception of discrimination for being muslim

The metric associated to this indicator is similar to the metric of swearing in the case of **I1**, as both metrics counts the number of tweets using a set of precise keywords (see Table 1). As in the latter metric, this value is calculated using the number of times a tweet contains a particular keyword, divided by the user's total number of tweets.

As it can be seen in Fig. 4, perception of discrimination for users of dataset D2 is heavily skewed to the left, that is, no tweets with discrimination keywords for almost every user. Again, a possible explanation to this observed effect is the wide range of written languages present in the dataset.

On the other hand, the ratio for dataset D1 is slightly higher than the ratio for dataset D3. This difference has been tested for statistical significance using a one-tailed Wilcoxon rank sum test with $W = 8551.5$, p-value < 0.05, even though both median are 0. This result emphasise the ability of this metric to distinguish the perception of discrimination for radicalised users.

4.5. Negative ideas about western society

As occur with the previous indicator, we used a set of keywords to measure the ratio of tweets expressing negative ideas about Western society for every user and dataset (see Fig. 5).

Regarding this metric we found the same results as with the discrimination indicator. The ratio of tweets with negative ideas from users in dataset D2 distributed around 0 as expected, since the metric is keyword based. In a similar way, it is more likely to get a higher ratio if the user belong to those considered as radicalised ($W = 8540$, p-value < 0.05), thus supporting the use of this metric as a numeric feature of **I4**.

4.6. Positive ideas about Jihadism

Finally, analogously to the metric of indicator **I4**, we measured the ratio of tweets per user that expressed positive ideas about Jihadism. Fig. 6 shows the results for the analysis of this metric. There is not much more to say about these results that have not been said previously: due it is a metric based on English keywords, it keeps failing at analysing tweets written in other languages (as Arabic), while there is a significant difference between the distribution of ratios for users in D1 and D3 ($W = 12\,265$, p-value < 0.05) which means that any user in risk of radicalisation is more willing to express positive ideas about Jihadism than a random user.

4.7. Analysis of the ratios

In order to discover the importance that each ratio has within the datasets, it have been studied their distribution over each dataset (see Fig. 7). As already explained, tweets from dataset D2 are written in many different languages, and those written in English represent a small part of the total. So it is not surprising that those keyword-based metrics have such low variability and their value is virtually zero. The only exception is the use of ellipsis, which seems to be a common practice in languages other than English as shown by the corresponding box-plot.

Focusing on the third dataset, which was made by the tweets from randomly selected users, there are three keyword-based metrics whose values are distributed on the low range. This makes

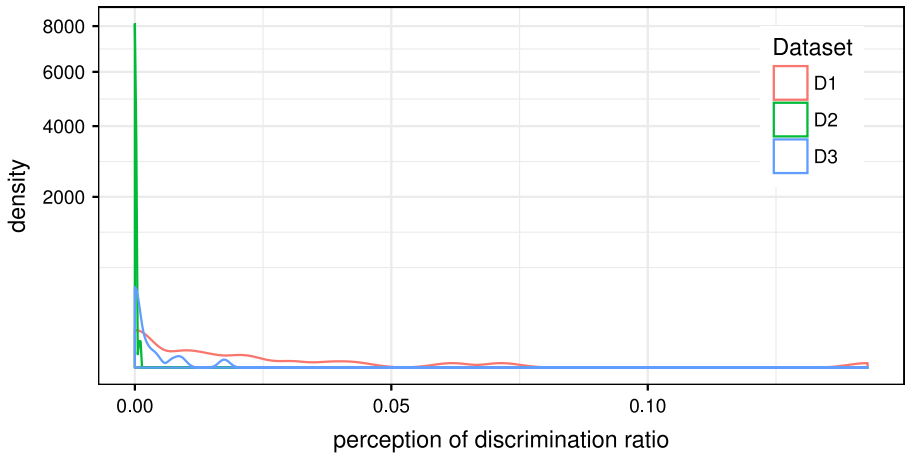

Fig. 4. Density distribution of the perception of discrimination ratio (exponential *y*-scale).

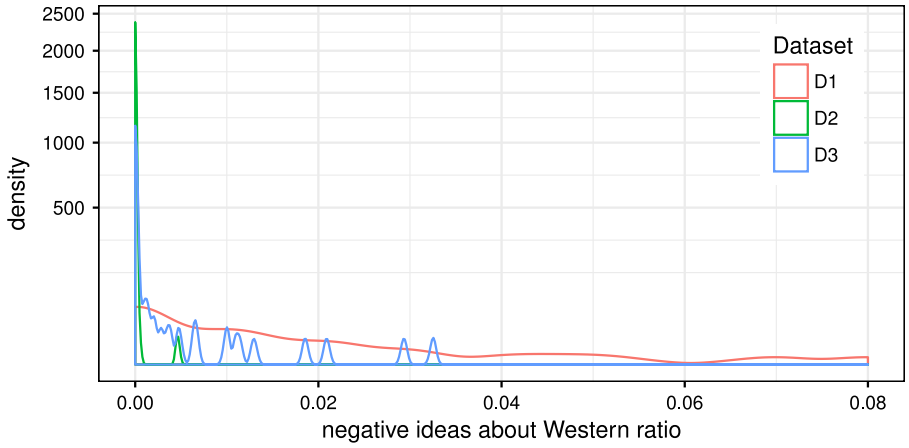

Fig. 5. Density distribution of the ratio of tweets expressing negative ideas about Western society (exponential *y*-scale).

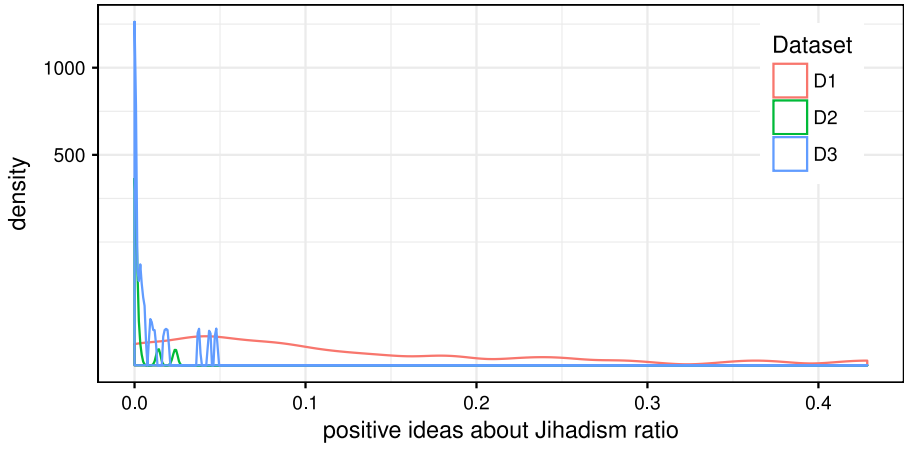

Fig. 6. Density distribution of the ratio of tweets expressing positive ideas about Jihadism (exponential *y*-scale).

sense as they are related to the keywords that express positive ideas about Jihadism, negative ideas about Western society and perception of discrimination, that is, they are related to very specific topics. On the other hand, ellipsis, swearing and negative content have more presence on the dataset, which suggests that is really common using and expressing them when writing messages on Social Networks.

In the case of the first dataset, the use of ellipsis follows the same behaviour as in the other datasets. As it was previously described, it is a fairly common procedure. The metric related to Jihadism keywords seems to have higher ratios than the other metrics when the user has been radicalised. This observation points out at the importance this metric should have when computing an aggregated metric to measure the risk of radicalisation.

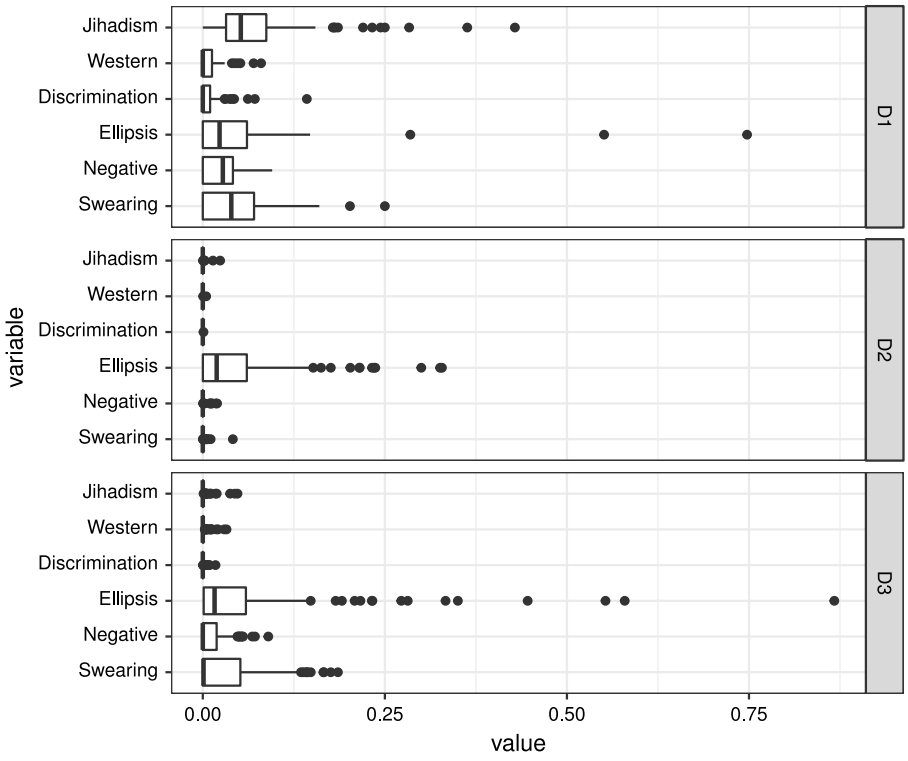

Fig. 7. Distribution of the different ratios (i.e. metrics) per dataset.

5. Conclusion and future works

This paper presents a study on a set of indicators devoted to risk radicalisation assessment, and their respective metrics used to quantify them. These indicators come from several experts psychologists on radicalisation, and are focused on the behaviour expressed in Social Networks. Furthermore, they can be easily computed in an automated way, which is an improvement over the current methods of risk assessment that are based on manual tools.

To study the performance of the aforementioned metrics, three datasets have been used: D1 that includes Twitter usernames and tweets from ISIS sympathisers, D2 that combines users and tweets flagged by Anonymous volunteers as radicalised people, and finally D3 that represents a random sample extracted from standard Twitter users. Precisely, we computed the metrics associated to the aforementioned indicators and then studied their density distribution. Then we compared the distributions of the datasets to discover if there is any statistically significant difference.

Regarding the ability of these metrics to highlight radicalised users, they perform generally well. We found statistical evidence that it is more likely to achieve a higher ratio of swearing, using words with negative connotations, perceiving discrimination and expressing positive and negative ideas about Jihadism and Western society, respectively, if the user is radicalised or in risk of radicalisation. Moreover, we found that radicalised users tend to write longer tweets than the rest, contrary to what was expected according to the introversion indicator.

On the other hand, we found that using ellipsis in the written text is not a relevant feature to use it as a metric for measuring introversion. This is a relevant result because this indicator has been traditionally considered as relevant in other areas as psychology [35,36]. Also, the performance of these keyword-based metrics were found to be extremely dependent of the language in which users wrote their tweets, so it is mandatory to increase the set of keywords with their corresponding translation in additional languages. This suggests that **I2** needs to be refocused in order to be as useful as the rest of the indicators.

Although the metrics studied in this paper have shown promising results as well as an adequate performance, they rely mainly on the messages and several set of keywords. It would be very interesting, as a possible future line of work, to define and analyse additional metrics based on the features and structure of complex networks, as Social Networks are considered of this kind.

The high dependency to the language exhibited by the indicators also lead us to think about integrating an ontology [37] to decouple those keywords expressed in a precise language and use language-agnostic concepts instead. Other possibility could be use directly a multilingual approach [38,39] for keywords and word processing based on the most relevant languages, as Arab, English, Russian, and French. This way, indicators will be able to operate on a wider range of messages, even capturing the use of slang or additional complex relationship between the expressed ideas.

Nevertheless, once the statistical suitability of the indicators have been tested, the next step is using these indicators as features for a Machine Learning algorithm that should be able to classify a Twitter user as being in risk of radicalisation by aggregating and combining the aforementioned indicators.

Acknowledgements

This work has been co-funded by the following research projects: EphemeCH (TIN2014-56494-C4-4-P) Spanish Ministry of Economy and Competitivity, under the European Regional Development Fund FEDER, and Justice Programme of the European Union (2014–2020) 723180 – RiskTrack – JUST-2015-JCOO-AG/JUST-2015-JCOO-AG-1. The contents of this publication are the sole responsibility of their authors and can in no way be taken to reflect the views of the European Commission. Finally, we would like to thank you to the reviewers and the Editor in Chief, for the different suggestions and comments made to this work.

References

[1] J. Scott, Social Network Analysis, Sage, 2012.
[2] S.P. Borgatti, A. Mehra, D.J. Brass, G. Labianca, Network analysis in the social sciences, Science (ISSN: 0036-8075) 323 (5916) (2009) 892–895.
[3] United Nations Office On Drugs and Crime, The Use of the Internet for Terrorist Purposes, Tech. Rep., Vienna, 2012.
[4] R. Thompson, Radicalization and the use of social media, J. Strateg. Secur. (ISSN: 1944-0464) 4 (4) (2011) 167–190.
[5] D.E. Pressman, J. Flockton, Calibrating risk for violent political extremists and terrorists: the VERA 2 structured assessment, British J. Forensic Pract. 14 (4) (2012) 237–251.
[6] D.E. Pressman, C. Ivan, 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 Radicalization in the Digital Era, first ed., IGI Global, Hershey, PA (USA), 2016, pp. 391–409 (Chapter. 19).
[7] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[8] D.J. Watts, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[9] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[10] G. Bello, H. Menéndez, D. Camacho, Using the clustering coefficient to guide a genetic-based communities finding algorithm, in: H. Yin, W. Wang, V. Rayward-Smith (Eds.), Intelligent Data Engineering and Automated Learning - IDEAL 2011: 12th International Conference, Norwich, UK, September 7-9, 2011. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN: 978-3-642-23878-9, 2011, pp. 160–169.
[11] S.E. Schaeffer, Graph clustering, Comput. Sci. Rev. (ISSN: 1574-0137) 1 (1) (2007) 27–64.
[12] G. Bello-Orgaz, H.D. Menéndez, D. Camacho, Adaptive k-means algorithm for overlapped graph clustering, Int. J. Neural Syst. 22 (05) (2012) 1250018.
[13] F.D. Malliaros, M. Vazirgiannis, Clustering and community detection in directed networks: A survey, Phys. Rep. (ISSN: 0370-1573) 533 (4) (2013) 95–142.
[14] A. Gonzalez-Pardo, J.J. Jung, D. Camacho, ACO-based clustering for Ego Network analysis, Future Gener. Comput. Syst. 66 (2017) 160–170.
[15] L. Egghe, R. Rousseau, Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science, Elsevier Science Publishers, 1990.
[16] R. Lara-Cabrera, C. Cotta, A.J. Fernández-Leiva, An analysis of the structure and evolution of the scientific collaboration network of computer intelligence in games, Physica A 395 (2014) 523–536.
[17] C. Cotta, J.J. Merelo, Where is evolutionary computation going? A temporal analysis of the EC community, Genet. Programm. Evolvable Mach. 8 (3) (2007) 239–253.
[18] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A (ISSN: 0378-4371) 311 (3–4) (2002) 590–614.
[19] J. Scott, Social network analysis, Sociology 22 (1) (1988) 109–127.
[20] M.S. Mizruchi, The American Corporate Network, 1904-1974, Vol. 138, Sage Publications, Inc, 1982.
[21] S.P. Borgatti, X. Li, On social network analysis in a supply chain context, J. Supply Chain Manag. (ISSN: 1745-493X) 45 (2) (2009) 5–22.
[22] T.W. Valente, K. Fujimoto, C.-P. Chou, D. Spruijt-Metz, Adolescent Affiliations and Adiposity: A Social Network Analysis of Friendships and Obesity, J. Adolesc. Health (ISSN: 1054-139X) 45 (2) (2009) 202–204.
[23] P. Wadhwa, M.P.S. Bhatia, Tracking on-line radicalization using investigative data mining, in: 2013 National Conference on Communications, NCC 2013. ISSN: 15503607.
[24] D. O'Callaghan, N. Prucha, D. Greene, M. Conway, J. Carthy, P. Cunningham, Online social media in the Syria conflict: Encompassing the extremes and the in-betweens, in: ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2014, pp. 409–416. ISBN: 9781479958771.
[25] E. Ferrara, W.-Q. Wang, O. Varol, A. Flammini, A. Galstyan, Predicting online extremism, content adopters, and interaction reciprocity, in: Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part II, Springer International Publishing, ISBN: 9783319478739, 2016, pp. 22–39. ISSN: 16113349.
[26] S. Agarwal, A. Sureka, (2015) Using KNN and SVM based one-class classifier for detecting online radicalization on Twitter, in: Distributed Computing and Internet Technology, 2015, pp. 431–442. ISBN: 9783319149769, ISSN: 16113349.
[27] M. Ashcroft, A. Fisher, L. Kaati, E. Omer, N. Prucha, Detecting jihadist messages on Twitter, in: Proceedings - 2015 European Intelligence and Security Informatics Conference, EISIC 2015, 2016, pp. 161–164.
[28] L. Kaati, E. Omer, N. Prucha, A. Shrestha, Detecting multipliers of jihadism on Twitter, in: Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015, 2016, pp. 954–960.
[29] L. Glowacki, A. Isakov, R.W. Wrangham, R. Mcdermott, J.H. Fowler, N.A. Christakis, Formation of raiding parties for intergroup violence is mediated by social network structure, Proc. Natl. Acad. Sci. (ISSN: 0027-8424) 113 (43) (2016) 12114–12119.
[30] D. Correa, A. Sureka, Solutions to detect and analyze online radicalization : A survey, Arxiv - Computers & Society V (January), 2013, pp. 1–30.
[31] I. Gilperez-Lopez, J. Torregrosa, M. Barhamgi, D. Camacho, An initial study on radicalization risk factors: Towards an assessment software tool, in: 2017 28th International Workshop on Database and Expert Systems Applications (DEXA), 2017, pp. 11–16, in press.
[32] R. Lara-Cabrera, A.G. Pardo, K. Benouaret, N. Faci, D. Benslimane, D. Camacho, Measuring the radicalisation risk in social networks, IEEE Access (ISSN: 2169-3536) 5 (2017) 10892–10900.
[33] R. Gladstone, Activist Links More Than 26,000 Twitter Accounts to ISIS, The New York Times, 31 March 2015. URL https://nyti.ms/2nA6A0B [Last Accessed: 04.04.17], 2015.
[34] G.A. Miller, WordNet: a lexical database for English, Commun. ACM 38 (11) (1995) 39–41.
[35] J. Merchant, The Syntax of Silence: Sluicing, Islands, and Identifying in Ellipsis, Oxford: Oxford University Press, 2001.
[36] C. Phillips, D. Parker, The psycholinguistics of ellipsis, Lingua (ISSN: 0024-3841) 151 (2014) 78–95. Structural Approaches to Ellipsis.
[37] N. Guarino, et al., Formal ontology and information systems, in: Proceedings of FOIS, vol. 98, 1998, pp. 81–97.
[38] M. Garcia, P. Gamallo, Yet another suite of multilingual NLP tools, in: J.-L. Sierra-Rodríguez, J.-P. Leal, A. Simões (Eds.), Languages, Applications and Technologies: 4th International Symposium, SLATE 2015, Madrid, Spain, June 18-19, 2015, Revised Selected Papers, Springer International Publishing, Cham, ISBN: 978-3-319-27653-3, 2015, pp. 65–75.
[39] A. Bérard, C. Servan, O. Pietquin, L. Besacier, Multivec: a multilingual and multilevel representation learning toolkit for nlp, in: The 10th Edition of the Language Resources and Evaluation Conference, LREC, 2016.

Raúl Lara-Cabrera obtained his M.Sc. and Ph.D. in Computer Science from the University of Málaga (UMA), Spain in 2013 and 2015 respectively. He is a research fellow at the Department of Ingeniería Informática of the Universidad Autónoma de Madrid, Spain. His main research areas involve computational intelligence, videogames and complex systems.

Antonio Gonzalez-Pardo is a Lecturer at Universidad Autónoma de Madrid. He received a Ph.D. in Computer Science (2014) from Universidad Autónoma de Madrid, a B.Sc. in Computer Science from Universidad Carlos III de Madrid (2009) and a M.Sc. in Computer Science from Universidad Autónoma de Madrid (2011). His main research interests are related to computational intelligence (genetic algorithms, PSO, SWARM intelligence... etc.), Multi-Agent Systems and Machine Learning Techniques. The application domains for his research are Constraint Satisfaction Problems (CSP), Complex Graph-based Problems, Optimisation problems and video Games.

David Camacho is currently working as Associate Professor in the Computer Science Department at Universidad Autonoma de Madrid (Spain) and Head of the Applied Intelligence & Data Analysis group. He received a Ph.D. in Computer Science (2001) from Universidad Carlos III de Madrid, and a B.S. in Physics (1994) from Universidad Complutense de Madrid. He has published more than 200 journals, books, and conference papers. His research interests includes Data Mining (Clustering), Evolutionary Computation (GA & GP), Multi-Agent Systems and Swarm Intelligence (Ant colonies), Automated Planning and Machine Learning, or Video games among others.