



Analysing the connectivity and communication of suicidal users on twitter



Gualtiero B. Colombo^{a,*}, Pete Burnap^a, Andrei Hodorog^a, Jonathan Scourfield^b

^aSchool of Computer Science and Informatics, Cardiff University, Queens Buildings, 5 The Parade, Cardiff, United Kingdom

^bSchool of Social Science, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, United Kingdom

ARTICLE INFO

Article history:

Available online 21 July 2015

Keywords:

Social media
Social network analysis
Twitter
Computational social science
Suicide

ABSTRACT

In this paper we aim to understand the connectivity and communication characteristics of Twitter users who post content subsequently classified by human annotators as containing possible suicidal intent or thinking, commonly referred to as suicidal ideation. We achieve this understanding by analysing the characteristics of their social networks. Starting from a set of human annotated Tweets we retrieved the authors' followers and friends lists, and identified users who retweeted the suicidal content. We subsequently built the social network graphs. Our results show a high degree of reciprocal connectivity between the authors of suicidal content when compared to other studies of Twitter users, suggesting a tightly-coupled virtual community. In addition, an analysis of the retweet graph has identified bridge nodes and hub nodes connecting users posting suicidal ideation with users who were not, thus suggesting a potential for information cascade and risk of a possible contagion effect. This is particularly emphasised by considering the combined graph merging friendship and retweeting links.

© 2015 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

It is recognised that media reporting about suicide cases has been associated with suicidal behaviour [1]. Concerns have been raised about how media communication may have an influence on suicidal ideation and cause a contagion effect among vulnerable subjects [2]. With the advent of open and massively popular social networking and microblogging Web sites, such as Facebook, Tumblr and Twitter (frequently referred to as social media), attention has focused on how these new modes of communication may become a new, highly interconnected forum for collective communication of suicidal ideation on a large scale. The demographic of online social networks is typically reported to be the younger generation [3,4] and thus teenagers and young adults are at particular risk. The risk of suicide contagion has been found to be especially high in adolescence and youth [5].

A limited number of studies have been published, reporting a positive correlation between suicide rates and the volume of social media posts that may be related to suicidal ideation and intent [6,7]. However, to date there is no study that is specifically focused on the connectivity and communication of suicidal ideation between users of

social media. Such a study could be important in the light of concern about the normalisation of suicidality and self-harm in social media. There is a small evidence base that suggests a connection between exposure to online self-harm- or suicide-related material and offline self-harming behaviour or suicidal ideation [3].

The research presented in this paper comprises an analysis of data collected from the microblogging website Twitter, the text of which has been classified as containing suicidal ideation by a crowdsourced team of human annotators. We study the connectivity characteristics between users and the propagation of suicidal content. To achieve this we have performed a social network analysis (SNA) of the connections of a specific subset of Twitter users who have been identified as posting content related to suicidal ideation. The SNA is applied to friend and follower connections of the subset of users, as well as investigating the potential content propagation by analysing the retweet graph of posts containing suicidal ideation. More specifically we are addressing the following research questions:

RQ1: With respect to the *friends-followers and mutual graphs* we focus on *measures of graph connectivity* to determine whether there is evidence of high connectivity between these specific type of 'suicidal' users, or whether these users are instead more isolated and exist within smaller social networks, as reported in [8,9]. Evidence that would allow us to partially answer this question is expected to be revealed by measurable network characteristics such as 'average node degree', 'graph density' and 'shortest path lengths'.

* Corresponding author. Tel.: +44 7976476926.

E-mail addresses: g.colombo@cs.cf.ac.uk, ColomboG@cardiff.ac.uk (G.B. Colombo), BurnapP@cardiff.ac.uk (P. Burnap), HodorogA@cardiff.ac.uk (A. Hodorog), Scourfield@cardiff.ac.uk (J. Scourfield).

RQ2: Regarding the *retweet graph*, we would expect traditional connectivity metrics to be less revealing as we do not have a complete network of all social ties (friends/followers) between retweeters. This is primarily because we only collected retweets for the sample set of 'suicidal users', due to the long time it would take to collect all users given the frequency/time limitations imposed by Twitter. Nevertheless, we can measure the *shortest path metrics*, which are a measure of information cascade. High values of average and maximal average shortest path imply greater propagation of information though the network. In addition, starting from an individual belonging to the set of 'suicidal' users, we can investigate if there is any evidence of social ties between these users and the Twitter users that have retweeted their posts. Evidence of this nature would allow us to gain insight into whether suicidal content is being restricted within the same community of friends and followers, or if it is propagating outside the user's social community into the wider network, where it could pose a risk of contagion.

The remainder of the paper is organised as follows. Section 2 describes the related work on this topic. Section 3 describes the data collection method. Sections 4 and 5 describe experiments used to measure connectivity and communication between suicidal users, and discuss the findings. Sections 6 and 7 draw conclusions from the study and identify possible ideas for future work.

2. Related work

A number of studies have recognised evidence that vulnerable subjects can be susceptible to the influence of news and reports of suicide in traditional mass media. The research literature on suicide clusters has supported the link between media reporting and suicide contagion and the impact of fictional and non fictional news stories of suicide [1,10]. There have also been recommendations for journalists about news reporting with particular emphasis on the language used in specific parts of a report, for example the headlines, and the differences between reports with national or local coverage [2].

In terms of the social network of groups of at risk subjects the majority of studies derive from medical research. For instance, in [11] the authors posed questions focused on social interactions in a poll of in-patients after a suicidal attempt, studying primarily the satisfaction level of social relationships reported by students and the unemployed. In [8] the authors conducted a similar study by investigating the relationship between friendships and suicidality among a larger sample of male and female adolescents in the US. Both studies came to the conclusion that an evaluation of the social network should be an integral part of the clinical investigation of suicidal related patients and form a basis for intervention. Furthermore, these studies provide motivation for the research presented in this paper.

However, only a small number of scientific articles have focused on the impact of social media communication. For example, in [6] the authors studied the potential of this new medium for predicting suicides by testing two social media variables (i.e. suicide-related weblog entries) over a period of three years, observing a positive correlation with suicide frequency. In [7] the authors conducted a study in the US on a dataset collected from Twitter using keywords and phrases related to suicide risk factors, filtered geographically by US state. Again they observed a positive correlation against national data of actual suicide rates.

Other studies have focused instead on the language used for the communication of suicidal thoughts, although they have primarily investigated other forms of written communication such as the classification of suicide notes (see [12,13]). This form of communication is typically more well-formed and less noisy than the type of short, informal language used in social media. Furthermore, the language was being expressed by people about to complete the act of suicide,

rather than those expressing thoughts of suicide. In [14] the authors report on depression-related language in Facebook¹. Facebook has less constraint on post length than Twitter, allowing more expressive thoughts to be posted; and we should not suggest that depression and suicidal ideation are synonymous, as they are not. Other recent studies have focused on depression and other mental health issues, highlighting the possible beneficial effects of social media communication [15–18].

More recently, there has been a more direct focus on the subjects potentially at risk of suicide, for example the Durkheim project² monitored the behavioural intent of a sample of US war veterans and analysed their social media posts on Twitter and Facebook to predict the risk of suicide ([19] and also [20]). However, none of these recent works looked specifically at the social network communication in terms of connectivity between users and propagation of suicidal ideation.

Social network connectivity has been studied by Hsiung [21] who reported the behaviour of an online mental health support group in reaction to a suicide case within the group. [22] reports how users who strongly express either positive or negative emotions heavily associate with each other, and [23] investigated the information contagion effect on a wider set of popular news stories in Twitter and Digg³. A systematic review of the research literature of Internet influences on the risk of self-harm or suicide, with particular focus on young people, is provided in [3].

Monitoring individual social media accounts to detect possible suicidal ideation is controversial territory, as evidenced by the recent withdrawal of the Samaritans Radar app in the UK⁴, but there is nonetheless potential to contribute to prevention as long as acceptability to social media users is thoroughly investigated. The research presented in this paper continues in this direction by focusing on Twitter as a case study for the analysis of connectivity and communication between people who post suicidal ideation. For the purposes of the paper we will refer to this subset of Twitter users as 'suicidal users'.

3. The collection of Twitter data

In order to collect and analyse suicidal communication posted to Twitter, we first needed to identify a set of terms that were likely to identify suicidal communication within text. To do this we initially collected text from Web forums via five Web sites^{5, 6, 7, 8, 9} either dedicated to discussion of suicidal thoughts and feelings or containing a large and easily identifiable body of such material. This resulted in 2000 anonymised forum posts that ranged in length from a few lines to several sentences and paragraphs. Each post was human annotated using the crowd-sourcing online service Crowdflower¹⁰. Human annotators were asked to identify content containing suicidal thoughts and feelings. Following the annotation we removed any annotations that were not agreed upon by at least four crowd-workers to be indicative of such emotion.

Term Frequency-Inverse Document Frequency (TF-IDF) analysis was applied to a each dataset (suicidal/non-suicidal). This process identified the most frequent terms in each dataset that are not present in the other, thus providing a ranked list of terms that are

¹ <https://en-gb.facebook.com>.

² <http://www.durkheimproject.org>.

³ <http://digg.com>.

⁴ <http://www.bbc.co.uk/news/technology-29962199>.

⁵ <http://www.experienceproject.com>.

⁶ <https://www.tumblr.com>.

⁷ <http://www.enotalone.com>.

⁸ <http://www.takethislife.com>.

⁹ <http://www.recoveryyourlife.com>.

¹⁰ <http://www.crowdflower.com>.

Table 1
TF-IDF listing of first 25 tri-grams and five-grams.

Trigrams		Fivegrams	
TF-IDF	3-gram	TF-IDF	5-gram
169.94	Want to die	32.819278	To take my own life
126.36	To kill myself	24.633562	Want to die right now
71.75	To commit suicide	22.590259	Have nothing to live for
68.18	Want to kill	19.691567	It's not worth it anymore
65.64	Can't live	19.691567	Don't want to live anymore
61.18	To end it	19.691567	Me want to kill myself
58.3	I'm tired of	19.691567	Myself hate my life hate
54.46	I hate myself	19.43643	Want to be here anymore
53.81	End it all	18.475171	Want it to be over
47.44	End my life	18.475171	Want it all to end
36.95	Take my own	18.475171	Wish could just fall asleep
33.89	Kill myself and	17.612125	Fall asleep and never wake
32.82	My death would	15.933278	Want to end it all
32.79	To live anymore	13.127711	Just really want to die
31.87	About killing myself	13.127711	Rather die its not worth
29.73	Kill myself i	13.127711	I'm sorry that im leaving
29.73	Never wake up	13.127711	Fuck trying to live normal
28.24	Killing myself i	13.127711	So why should continue living
26.26	Stop the pain	13.127711	Don't want to live defeated
26.26	Kill myself right	13.127711	To commit suicide within few
25.89	Thoughts of suicide	13.127711	And pain anymore just can
25.89	Point in living	13.127711	Put an end to this
24.63	Worth it anymore	13.127711	Been self harming for years
24.3	Have nothing to	13.127711	Bad really am worthless what
21.86	Wanted to die	13.127711	Life is this miserable just

more likely to be suicidal than not. In this study, we considered terms as n -grams of up to five tokens in length. To further penalise common phrases and words that appear in both suicidal and non-suicidal contexts, while prioritising terms belonging exclusively to the former dataset, TF-IDF was applied by considering the posts classified as non-suicidal as distinct documents, whereas those including suicidal intent were aggregated into a unique document. Examples of the most relevant trigrams and five grams produced by the TF-IDF procedure are given in Table 1.

Because of the significant number of irrelevant terms that would not logically be useful as search keywords for the Twitter data collection, the TF-IDF lists were subject to further examination by two experts in the suicide field leading to a list of 62 key words and phrases used to collect suicidal communication from Twitter, as shown Table 2.

Illustrative examples are *asleep and never wake*, *don't want to exist* and *kill myself*. These search terms were then used to collect

data from Twitter via the Twitter Application Programming Interface (API)¹¹.

Twitter is a micro-blogging site with 255 million active users worldwide posting an estimate of over 500 million Tweets per day¹² on an open and accessible basis. This makes Twitter a suitable source of data for a study into connectivity and propagation of suicidal ideation, but also results in an extremely noisy environment, where posts cover a large variety of topics. As a consequence, the data retrieved are required to be pre-filtered in order to consider a sufficient number of posts that can be classified as containing suicidal ideation.

Data were collected from Twitter for a six-week period starting on the 1st February 2014, resulting in over four million posts. As a parallel activity, we monitored traditional media over the same period to identify the names of suicidal cases of young people in England (focusing on the teenage range of 11–18 years old) and then searched and retrieved data from Twitter containing the name and surname of the deceased. Using the 'names' dataset, 2 expert suicide researchers discussed the features of the Tweets and derived a coding frame concerning not only suicidal thinking and ideation (also including expressions of total despair, even if suicide is not explicitly mentioned) but also memorials, campaigning, information and support, and news reporting. The following seven-class coding frame was developed by these researchers to capture the best representation of how people generally communicate on the topic of suicide.

- 1: Evidence of possible suicidal intent
- 2: Campaigning (i.e. petitions etc.)
- 3: Flippant reference to suicide
- 4: Information or support
- 5: Memorial or condolence
- 6: Reporting news of someones suicide (not bombing)
- 7: None of the above

We then extracted a random sample of 1000 tweets from the 4 million collected over a six-week period and repeated the human annotation task using the same crowdsourcing service¹³, this time asking crowd-workers to classify Tweets into a number of suicide related categories. The reason for selecting a sample of 1000 is that human annotation is a manual and time-intensive task. Similar research into the classification of emotive texts using a human annotated gold-standard has typically used a sample of 1000 to good effect [24–27].

¹¹ <https://dev.twitter.com/overview/api>.

¹² <http://www.internetlivestats.com/twitter-statistics/trend>.

¹³ <http://www.crowdfunder.com>.

Table 2
Keywords and phrases search terms.

Asleep and never wake	Just want to sleep forever	Take my own life
Can't do this anymore	Kill myself	Thoughts of suicide
Could just fall asleep	Killing myself	Tired of being alone
Die in my sleep	Life is so meaningless	Tired of being lonely
Don't want to be here	Life is too hard	To end this nightmare
Don't want to exist	Life is worthless	To hurt myself
Don't want to go on	My death would	To live anymore
Don't want to live	My life consists of nothing	Want it to be over
Don't want to try anymore	My life is pointless	Want to be alive anymore
Don't want to wake up	My life is this miserable	Want to be around anymore
End it all	My life isn't worth	Want to be dead
End my life	Not want to be alive	Want to be gone
End this pain	Nothing to live for	Want to be here anymore
Ending it all	Point in living	Want to die
Hate my life	Put an end to this	Want to disappear
Hate myself	Ready to die	Want to end it
I'm drowning	Really need to die	Wanted to die
I'm leaving now	Stop the pain	Wanting to kill yourself and
I'm worthless	Suicidal	What is wrong with me
Isn't worth living	Suicide	Why should I continue living
Just want to give up	Take it anymore	

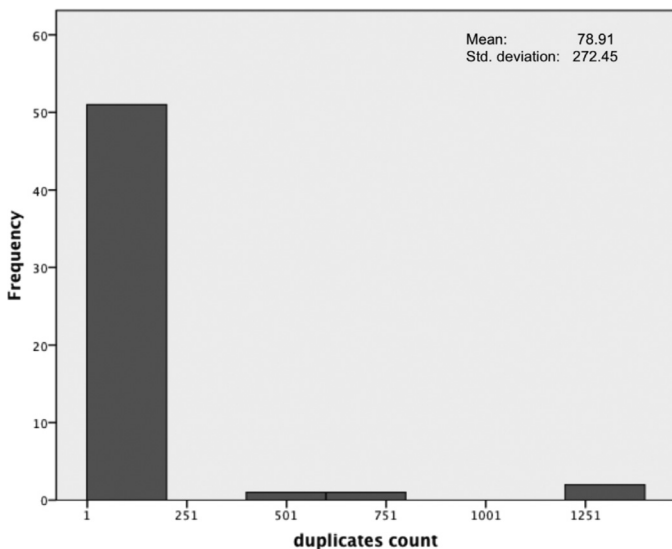


Fig. 1. Distribution of duplicates of the initial set over 71 suicide related Tweets.

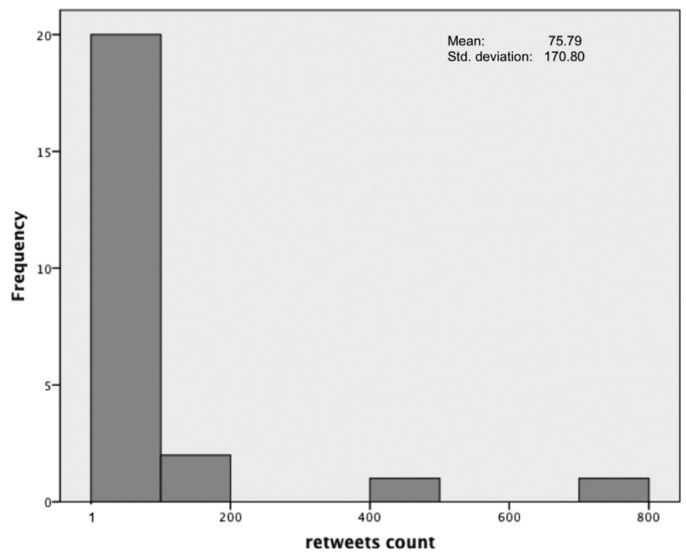


Fig. 2. Distribution of retweets over the complete set of 4543 suicide related Tweets.

Our main interest was in the first class of posts containing evidence of possible suicidal intent. As may be expected, this particular type of content is present in Twitter only in a small minority of posts. Following the second human annotation task we removed all Tweets that had less than 75% agreement among crowd-workers and obtained a set of 71 posts classified into this first class (11.8% of a total of 601 with at least 75% agreement among human annotators).

To extend the datasets of Tweets on which to perform our analysis, we also considered any duplicates (Tweets with exactly the same text) of the initial set of 71 that were contained into the whole six-week collection of pre-filtered Tweets. This resulted in a total of 4543 posts that constitute our final dataset of Tweets (human) classified as containing possible evidence of suicidal intent. The distribution of the duplicates is shown in Fig. 1 showing how the majority of Tweets included into the initial set had only a small number (in the order of units) of exact copies of the same text out of the whole datasets, while only a handful of them had more than a few hundred. We define the whole set of authors of these posts as the set S (or 'suicidal' set) throughout the paper, for a total of 3535 Twitter users posting this type of content.

Finally, for each Tweet in the resulting set of 4543, we collected all retweets contained in the whole six week dataset. We identified retweets following a pattern recognition technique that extracted them out of the whole six weeks collection as any post matching the following format: 'RT ' + space + '@screenname' + space + ':' + 'Tweet text' + 'some more text (if any)'. This resulted in 2365 retweets, for which Fig. 2 illustrates the distribution, showing long-tail characteristics where the majority of tweets have very few retweets, but a small number of them have been widely propagated.

4. The friends and followers distributions - measures of connectivity

For each of the authors of the 4543 Tweets classified as containing evidence of possible suicidal intent we retrieved Twitter profile information pertaining to the lists of *followers* and *friends* (users followed) so that we could identify measures of connectivity between this type of user. This resulted in two very large sets of 2,376,559 followers and 1,600,498 friends for a list of 3535 distinct authors.

The graph of followers is a directed graph (with the out-going edges meaning a *is followed by* relation). Our data show an average number of followers of 528 per user, which is more than double

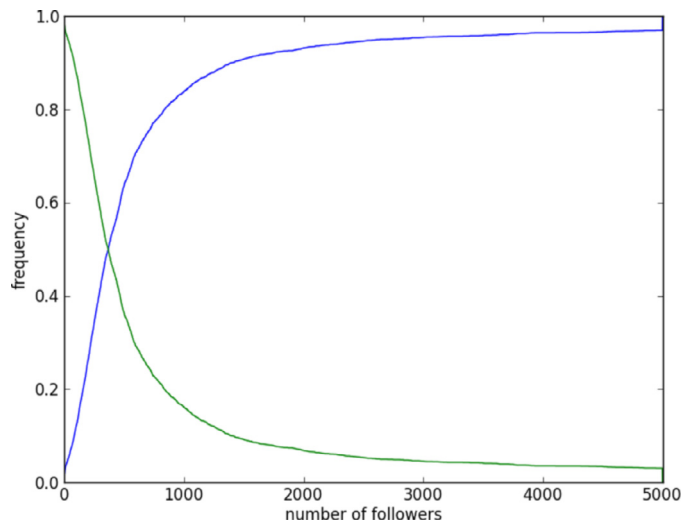


Fig. 3. Cumulative (blue) and survival (green) distributions of followers over the complete set of 4543 Tweets containing suicidal intent. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

the Twitter average of 208.¹⁴ This would suggest a higher than average level of 'social capital' within the 'suicidal' users in the set S , where 'social capital' is a measure of how many people are likely to receive information from the user. Celebrities and politicians typically have high levels of Twitter social capital. The survival (1-cumulative) distribution of followers mirrors the characteristics reported in other studies of follower distributions [28,29], as visible in Fig. 3.

We also computed the distribution of 'friends' (users followed) and a 'mutual' list of users that reciprocally follow each other. Having a 'following' relationship with many users who post suicidal content could be interpreted as being a 'consumer' of such content, while a mutual connection could suggest mutual interest in sending and receiving content. The resulting averages per user were 372 and 313 respectively for 'friendship' and 'mutual' links with statistical distributions similar in their long-tail shape to the one obtained for the followers lists (here omitted for reasons of space).

¹⁴ <http://expandeddrablings.com/march-2013-twitter-stats/>.

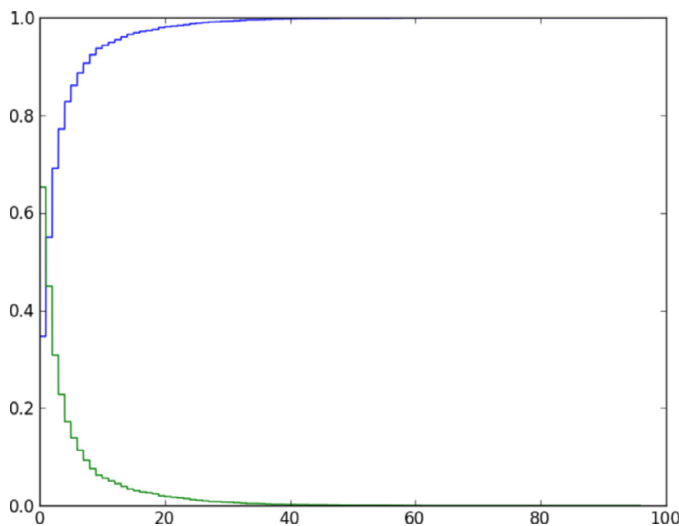


Fig. 4. Cumulative (blue) and survival (green) distributions of followers belonging to the set S of 'suicidal' users (over the complete set of 4543 suicide related Tweets). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

The list of friends and followers presented so far refers to the aggregate of all the friends/followers returned by the Twitter API¹⁵ for each of the set of 'suicidal' users. Note that the users in these lists were not necessarily expected to belong to the initial set S . However, we were interested in the degree to which this occurs, to establish if there are mutual friendship relationships between users posting suicidal content. This can provide evidence of communities existing around this topic. Fig. 4 confirms that there is indeed a level of reciprocal friendship between users posting suicidal ideation, as evidenced by the survival (the reciprocal of cumulative) distribution. Although it stills follows a long-tail distribution, with the vast majority of users having a small number of links, a notable percentage of users (about 20%) appear to have links with other 'suicidal' users.

4.1. Graph representation of friends and followers

Following our identification of some level of connectivity between suicidal users, we proceeded to build graph representations of followers, friends and mutual friends. Here nodes represent users that belong exclusively to the set S of 3535 'suicidal' users and edges the 'follow', 'friendship' (directed) and 'mutual' (undirected) links between pairs of users included in this particular class. Figs. 5–7 shows the graph representation of the followers graph resulting in 833 nodes and 273 edges, having here discarded users that did not have any follower connection within S .

Fig. 5 shows a very sparse graph with many small disconnected sub-graphs visible in the outer circle. However, also visible is a core of nodes that appear connected via a follower relationship. The core of this network is expanded in Fig. 6. In this figure the nodes' sizes and colours follow a scale according to their degree representing the *is followed by* relation. The nodes range from red to blue, where red nodes have many followers (more followers = larger node size) and blue nodes have less or no followers but are following the most people. Similarly red edges represent the *is followed by* relationship and blue edges represent *follows*. Here we can observe the presence of large red nodes that have a function of 'hubs' in the graph being connected with ('followed by') several other nodes (see also the graph detail in Fig. 6). These nodes could be seen as influential users within

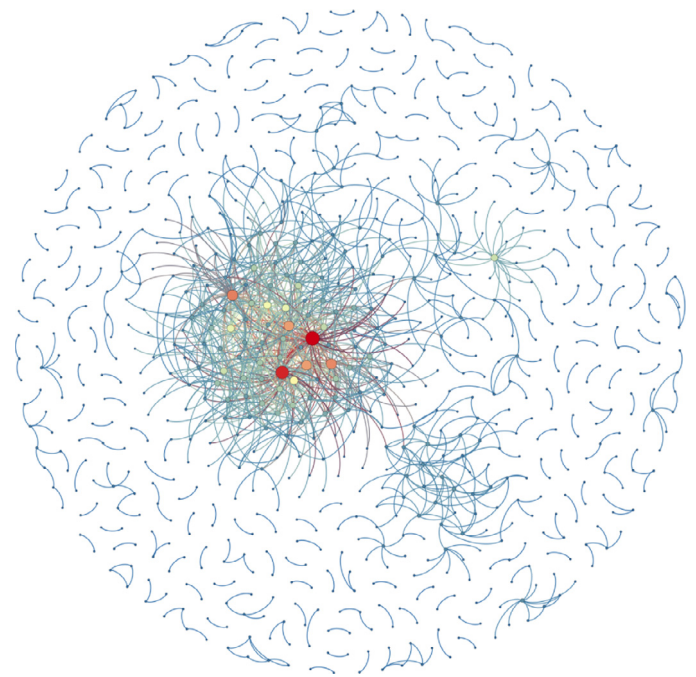


Fig. 5. Graph representation of the followers graph of users $\in S$.

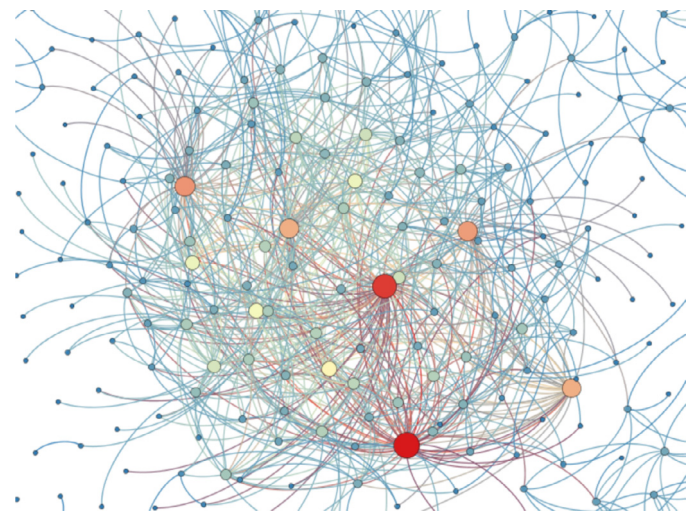


Fig. 6. Particular of the core sub-graph of the followers graph among users $\in S$.

the community, having high social capital and the potential to communicate with a wide range of other suicidal users.

Fig. 7 shows a 'close up' of one of these hubs. Note that the large size of the node implies the existence of a considerably large set of followers. Moreover, we can observe that this followers set includes other red and orange nodes of considerable size themselves, that in turn have a number of their own followers. This can produce high potential for the spread (cascade) of information over the network.

Nodes in between the red and blue range (in the order of orange, light yellow and light green nodes) can be seen instead as intermediate nodes having both followers and following other nodes (in different proportions following the colour order). They then form potential communication bridges among different communities (see 6). Connecting two communities is therefore likely to support contagion between graphs.

Table 3 summarises a number of metrics for the following three graphs of followers, friends and mutual connections. These results provide the statistics for:

¹⁵ <https://dev.twitter.com/overview/api>.

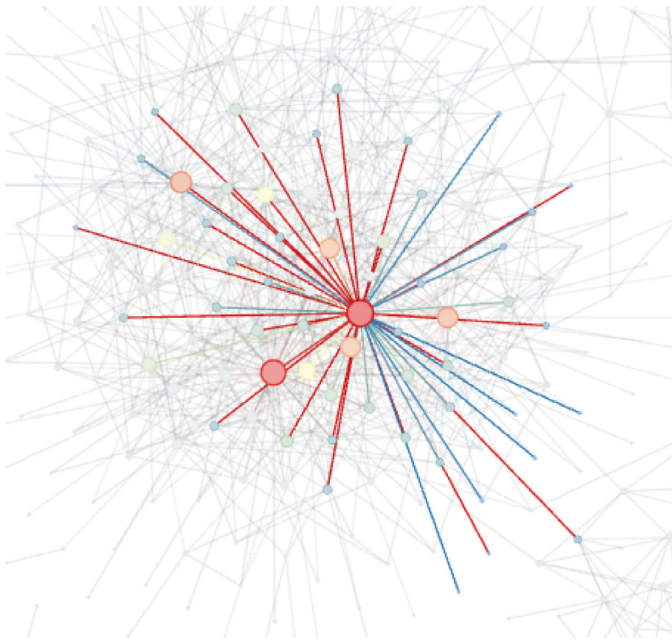


Fig. 7. Particular of a hub node in the followers graph of users $\in S$.

Table 3

Graph metrics for followers, friends and mutuals of 'suicidal' users.

Metric	Foll.	Fr.	Mut.
Nodes	833	863	607
Edges	1273	1423	958
Density	3.7E-03	3.8E-03	5.2E-03
Conn	172	161	92
LCC	377	435	352
Avg. Deg.	3.06	3.30	3.16
Max. Deg.	53	59	53
Avg Clust.	0.063	0.082	0.062
Triang.}	1869	3150	1401
Trans.	0.14	0.18	0.13
Avg. sh.	4.79	4.99	4.93
Diameter	14	16	15

- *Number of nodes*: The number of vertices in the graph.
- *Number of edges*: The number of links connecting pairs of vertices.
- *Graph density*: The ratio between the number of edges in the graph and the total number of possible edges.
- *Average graph degree*: For each vertex the degree is calculated as the number of links that end in that vertex. For the directed graph such as the followers and friends we have calculated the out degree (number of outgoing edges) representing respectively the 'is followed by' and 'is following' relations. The *average degree* computes the average of the degree values over all network nodes.
- *Max graph degree*: The maximum value of the nodes degree over all graph vertices.
- *Number of connected components*: The number of sub-graphs for which any two vertices are connected to each other by edges.
- *Largest connected component (LCC)*: The maximum size (number of nodes) of a connected sub-graph.
- *Average clustering coefficient*: Firstly we calculate the clustering coefficient for each node as the probability that two randomly chosen distinct neighbours of the given node are connected. This is also referred to as the *local clustering coefficient* for a node. Then we average these values over all network nodes.
- *Number of triangles*: Number of triples of nodes all connected pairwise by an edge.

- *Transitivity*. This is another *global measure of clustering* and is proportional to the ratio between the total number of triangles and the number of connected triples of vertices (groups of three nodes with at least two edges connecting pairs of them).
- *Average shortest path*. We firstly defined the shortest path length between two nodes as the number of edges (hops) that we need to travel through to connect one to the other. This is equal to one when nodes are linked directly by an edge, and higher if there are any intermediate nodes and edges that connect the two extremes represented by the given pair. We then compute the shortest path value when more than one of such paths exist. For a node the *average shortest path* is then defined as the average of the shortest path values between the given node and all others in the graph.
- *Maximum shortest path*. The maximum value of the shortest path calculated over all pairs of vertices in the graph. This is also referred to as the *diameter* of the graph.

A mathematical formulation of all the metrics listed above can be found in [30]. All above metrics aim to measure how nodes are linked to each other and, consequently, how they can potentially disseminate content from a node to its neighbouring nodes (friends, followers), and from them to their own neighbours and so on. More specifically:

- *Degree (avg, max)* and *density* are essentially measures of graph connectivity in terms of links/relations between nodes. This, in terms of follower/following degrees, means that users can directly consume (see, read) the content posted by other users.
- *Average clustering coefficient* and *transitivity* are both *clustering* metrics that measure how some of the nodes can form dense groups in which each element has strong connections with the others. As a consequence, each piece of information posted by one of these nodes can rapidly spread within the groups but disseminates outside the group with more difficulty. Note that if the graph nodes were all connected to each other we would have only one big cluster (this is also expressed by high *density* values that can then be seen as a measure of 'global clustering'). However, usually (as in our graphs) a number of finite clusters are visible, normally having weak connections between each other (*weak ties*). If no connections at all exist between clusters we would define them as disconnected components. When many nodes are included in one of these clusters the average clustering degree values become higher - even if the graph appears composed by many distinct clusters.
- *Shortest paths* metrics are a direct measure of how information travels throughout the network, following paths represented by links between a node and his neighbours, between them and their own networks, and so on. The greater the length of the shortest paths from a node to all others in the graph (and so their average), the easier the information can travel from a given node and spread over the network. The flow of information spreads with increasing difficulty beyond the edge of the connected components and clusters of nodes. However, as observed earlier, clusters could still be connected by a small number of links (*weak ties* [31]) that act then as *bridges* between cluster pairs and allow information to spread from a vertex to the others leading to a possible *contagion effect* (this is reflected by greater values of each node shortest paths to all other network nodes).

From the values in Table 3 we can observe that the graphs representing the followers and friends networks are very similar, with the latter having slightly greater degrees and clustering indexes (e.g. average degree, average clustering). This is also reflected in the higher number of triangles and greater transitivity, meaning a slightly more connected graph.

Secondly, we can observe that the graph built with mutually reciprocated links shows very similar values for the majority of the

Table 4
Graph metrics for baseline Twitter networks.

Metric	k1	k2	k3
Nodes	465,017	52.5 m	41.6 m
Edges	834,797	1.9b	1.4b
Density	3.2E-06	1.4E-07	1.6E-07
Conn	–	–	–
LCC	465,017	–	–
Avg. Deg.	3.59	74.68	70.51
Max. Deg.	678	3.6 m	3.1 m
Avg Clust.	0.061	–	–
Triang.	38,389	55.4b	34.8b
Trans.	–	–	–
Avg. sh.	4.59	–	–
Diameter	8	18	23

metrics of connectivity, such as maximum and average node degree, clustering coefficients, average shortest path, diameter, and even higher graph density (see Table 3).

For baseline comparison of social network metrics we refer to three datasets publicly available from the website Konect [30] (the Koblenz Network Collection), which provides large network datasets for scientific research. We will refer to these as ‘baseline network metrics’. In Table 4 we provide network metrics (when available) for the three following datasets of different sizes (all representing Twitter follower networks):

- k1 - Twitter (ICWSM)*: directed network containing information about who follows whom on Twitter.
- k2 - Twitter (MPI)*: asymmetric network containing Twitter ‘follow’ data based on a snapshot taken in 2009.
- k3 - Twitter (WWW)*: follower network from Twitter, containing 1.4 billion directed ‘follow’ edges between 41 million Twitter users.

Although Twitter networks of different size and nature inevitably show different characteristics, the graphs of ‘followers’, ‘friends’ and ‘mutuals’ present a density of three degrees of magnitude greater than the benchmark datasets ‘k1’, ‘k2’ and ‘k3’ used for comparison (in the order of E-03 instead of E-06). These values further drop with the increasing size of the graphs, thus suggesting that, although of generally low density, the level of interconnectivity between ‘suicidal’ users may be greater than that in these baseline networks. The opposite happens for the average degrees, suggesting instead that these users are more isolated from other users than in the baseline networks. However, the network of ‘suicidal’ users is actually relatively small compared to the baseline networks and our results show that the measures that express connectivity, such as the average degree and the average clustering coefficient, are comparable between our values and those of the smallest Konect graph *k1*.

A further published work also provides an analysis of the Twitter ‘follow’ graph, taking a snapshot from the second half of 2012, by defining four different networks of different size [28]. The degree of connectivity is here very similar to our results, with the range of average degrees varying from 2.83 to 3.34 for the follower graph, from 3.56 to 4.03 for the friend graph, and from 2.59 to 2.83 for the graph representing ‘mutual’ links. The distribution of clustering coefficients is also comparable with our findings (0.19 for nodes of degree 20). This again suggests that the connectivity within the suicidal user set is similar to the generic Twitter network connectivity. This study also reports an average path length of 4.17 for the ‘mutual’ graph and 4.05 for the directed graph of followers for the networks, while we obtain values of 4.79 for the followers and 4.93 for the ‘mutual’ links, providing further evidence of a connectivity among suicidal users which is comparable to that of generic Twitter users.

Moreover, the authors report that 42% of edges in the ‘follow’ graph are reciprocated, whereas our graphs return much higher percentages with 75 of the ‘follow’ links also having ‘friendship’ links

Table 5
Graph metrics for retweets.

Metric	Re-tw.	Re-tw+Fr.
Nodes	3209	3866
Edges	2211	3469
Density	4.3E-04	4.6E-04
LCC	138	827
Conn	1002	1023
Avg. Deg.	1.38	1.79
Max. Deg.	44	69
Avg Clust.	9.4E-03	0.013
Triang.	9	1878
Trans.	1.4E-03	0.08
Avg. sh.	5.05	5.43
Diameter	13	15

between the two nodes. This result is in line with other recent studies that have identified in large networks the presence of sub-communities of members highly associated to each other. Furthermore, the same studies suggest this may be correlated to the high emotional state of these members, such is the case of our network of ‘suicidal’ users that forms itself a sub-community of the much larger Twitter network.

Nevertheless, the fact of recording a degree of connectivity comparable to that of other snapshots of more generic Twitter users in terms of social network metrics (apart some predictable differences from the largest graphs of several million of users) is an important result itself. In fact, our network is formed exclusively by users belonging to the ‘suicidal’ set (having discarded any ‘follow’ and ‘friendship’ links with nodes outside this given set) and has been generated by only considering the authors of a very small sample of distinct Twitter posts (originally less than one hundred annotated as ‘suicidal’ and then expanded by considering their duplicates in the collected data). As a consequence no particularly significant degree of connectivity was expected among this resulting group of users.

5. The retweet graph - measures of communication

This section analyses the graph of retweets, built by looping through *S* and identifying which users have retweeted posts containing suicidal ideation. This has the effect of further propagating this type of content and may increase the risk of contagion. The retweet graph is a directed graph where the direction of the arrows means ‘has retweeted’. A summary of graph metrics related to the retweet graph is given in Table 5. Only a relatively small percentage of our initial set of users have been retweeted ($1036/3,535 = 29\%$), as visualised in Fig. 2 suggesting a long-tail distribution. This also means that only 32% of the nodes in the retweet graph are from the initial set *S* of ‘suicidal’ users.

In Table 5 we can observe very low values for all the connectivity metrics (such as degree, clustering, and a much higher number of disconnected components) in comparison with those obtained from the follower and friend graphs. This is, however, a consequence of the fact that we focused intentionally only on posts included in the annotated set of human classified suicidal tweets, thus only considering retweets of this particular group of users without incorporating those who have not been identified as posting suicidal ideation. As a result, the retweet graph does not include any edges without at least one end included in the set *S*.

Therefore, our collection only explored retweet links going *one-hop* away from our initial set of users and so missing out potential triangles among triads of nodes when these were not all included in our given set (as in the majority of cases). This resulted in a reduction in the indexes of transitivity and clustering, whereas the average degree still achieves a third of the values obtained for the followers and friends networks.

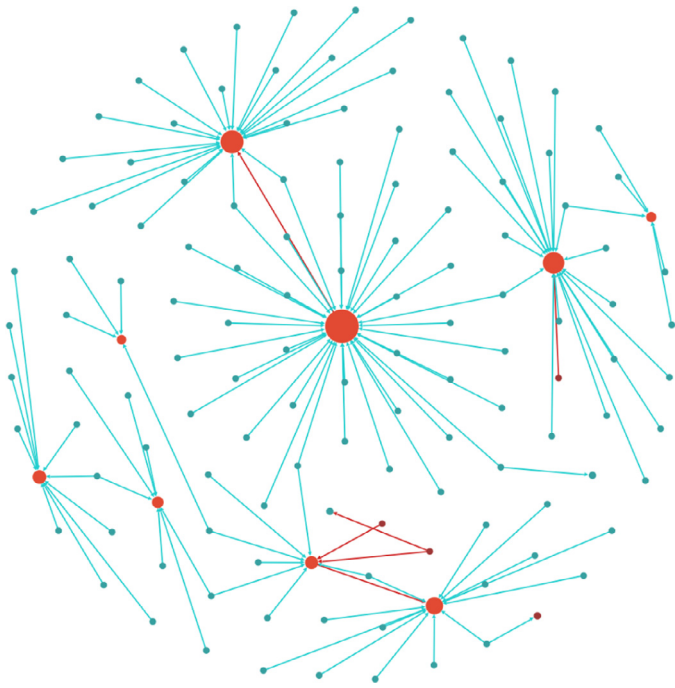


Fig. 8. Particular of hub and bridges in the retweet Graph nodes $\in S$ (red) - nodes $\notin S$ (blue) edges $\in S$ (red) - edges $\notin S$ (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

However, from the analysis of metrics other than connectivity indexes we can observe interesting properties. [32] reports an extensive study of a large datasets of a 2009 snapshot of the Twitter graph analysing hundreds of thousand of users and their retweets. It concludes that, even if the retweet graph shows the same scale-free characteristics, it presents a higher degree of connectivity than typical online networks. In particular the authors observed larger connected components and higher clustering coefficients (greater than in the follower graph) resulting in a closer behaviour to real-world networks in terms of content dissemination. The latter property is captured by the values of the average shortest path (4.8) and diameter of the graph (8.5). Similar results are also reported in [33] that analysed over four thousand retweet groups (for a total of about 26,000 Tweets) collected over the year 2011. The authors obtained a maximum longest shorter path over all groups of 9 edges (although the average shortest path was much lower and only equal to 2). Our results, presented in Table 5, show higher values of both the diameter (maximum shortest path of 13/15) and average shortest path (between 5 and 5.5). This finding suggests a greater spread of suicidal ideation content than that observed for typical Twitter content in the comparable studies.

The average shortest path in our retweet graph is also in line with that reported in a public Konect dataset (5.45) which represents a much larger Twitter network of online interactions ('mentions'), with three million nodes and over ten million edges [30]. This provides further evidence that the 'suicidal' user network S presents properties similar to large scale communication networks, thus suggesting a high level of propagation of such content within the virtual community and some potential for information spread (and a possible contagion effect).

The propagation of information can also be explained by looking at particulars of the retweet graph (see Fig. 8), which appears as highly disconnected (very sparse with over one thousand connected components) with most of the users only connected in small size disconnected sub-graphs usually formed by small hubs with at the centre a node $\in S$ ('suicidal' nodes) and at the edges a small group of nodes external to S . However, the relatively high shortest path val-

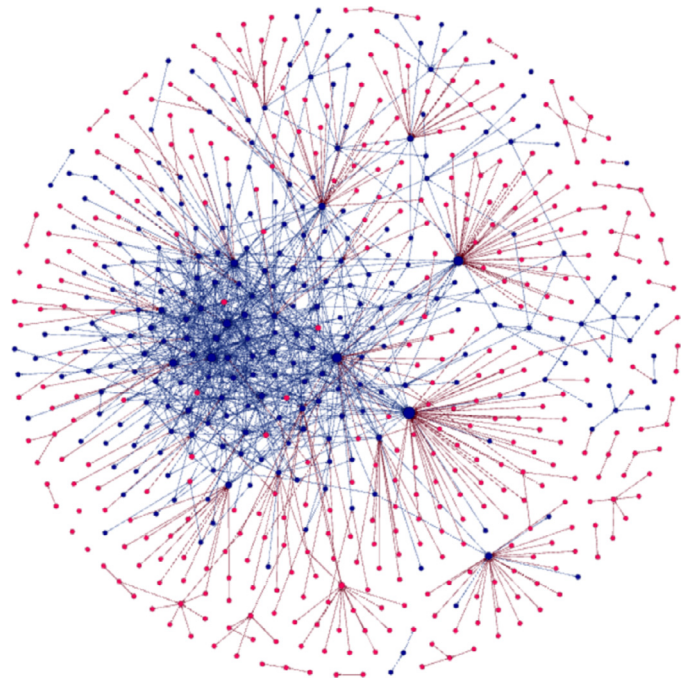


Fig. 9. Combined Graph of retweets and 'follow' links 'follow' edges (blue) - retweet edges (red) nodes $\in S$ (blue) - nodes $\notin S$ (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

ues suggest the existence of weak links/bridges that connect together different hubs.

Even if not numerous, these weak links and bridges do exist in our graph, as observable from Fig. 8. Here nodes belonging to S are represented in red while 'external' nodes are coloured in blue. The size of the user/node is proportional to the number of retweets for original 'suicidal' tweets posted by that user. We can observe a number of 'hubs' where the centre of the hub is a user that posted suicidal content, which has subsequently been retweeted a number of times, since these nodes appear of a considerable size. Surrounding the hub are retweeters who are (in the majority of cases) external nodes (not in S), thus allowing content dissemination outside our initial set of suicidal users. Once again, this provides evidence of a possible contagion effect. Also note the importance of a number of 'bridge nodes' that have retweeted (and so linked together) pairs of different hubs. In Fig. 8, edges represent the relation 'has retweeted'. Edges between nodes external to S and internal ones are coloured in blue and appear as the large majority, whereas only few links (in red) present both ends belonging to set S (red nodes).

This is also in line with recent studies, see [34], that emphasise the importance of 'weak-links' within the Twitter network for the dissemination and sharing of content.

5.1. Combining friendship and retweet links

As a final step, we merged the two graphs of followers and retweeters, thus adding 'friendship' edges to nodes in the retweet graph as well as adding users from S that had 'follow' links but have not retweeted each other. The purpose of this is to identify levels of propagation between suicidal users.

The network metrics for this 'combined' graph are given in the second column of Table 5. Here we can observe that the size of the larger connected component, the number of edges, the degree, and clustering indexes have all increased, suggesting a very dense and connected community with high volumes of propagation.

This is visible in Fig. 9 that also visualises how these links are related to each other, since 'friendship' means potentially

consuming a user's content while 'retweeting' is a clearer index of content already consumed. In particular we are interested in retweets that are made by users that are not already part of the 'suicidal' set S (blue indicates nodes $\in S$). From the Figure we can observe how these retweets (represented as red edges) are primarily located on the outer circle and produced by retweeting components of small size (mostly pairs) that appear in isolation from the rest of the network.

This is further supported by the shortest path metric values in Table 5 not being affected to a significant extent by the addition of the 'friendship' links. In fact, although degree and clustering indexes increase because of the addition of them, the shortest paths appear not to shorten (but instead slightly increase). A shorter length may be expected if the majority of retweets were done by users within the suicidal set that are already connected by 'friendship' links. Note that this result is in line with other recent studies, such as [28] that reports longer shortest path values for larger Twitter graphs and is in contradiction with what has been observed for other social networks, suggesting that the average path length should instead decrease with the size of the graph [35].

From this figure we can again observe how, beside a dense network of friendship links among 'suicidal' users in the inner part of the graph (blue edges), retweeting of suicidal content is performed by users who are not connected and do not belong to S (red edges). This suggests that the propagation of suicidal ideation may not occur among 'suicidal' users but instead the dissemination of this specific type of content could be enacted by users who are not directly connected to them.

6. Conclusion

In this paper we have analysed the graph characteristics of a set of 3535 Twitter users who have posted content that human annotators agreed should be classified as containing evidence of suicidal thinking. For the purposes of the research, we refer to these users as 'suicidal users'.

We conducted a range of social network analysis experiments by analysing the social graphs derived by identifying the followers, friends, mutual friends (where both users follow each other), and retweets of suicidal users. Each node in the social graphs belonged to the given set of 'suicidal' users. A number of significant characteristics and properties have been observed by analysing these graphs.

With respect to connectivity, the friends and followers graphs of suicidal users did not present major differences in terms of social network metrics when compared to other literature reporting Twitter snapshots of more generic users (apart from predictable differences from very large networks of millions of users). However, our results showed that while the average user connectivity metrics appear similar to baseline networks, the reciprocity of either follower/following relationships or 'mutual' links between suicidal users is significantly higher (up to 73% as opposed to 42% in other studies), suggesting a more tightly-bound community than non-suicidal networks.

From the investigation into communication, our study found that the values of the average shortest path of retweets of suicidal content were higher than in previous studies that reported on general retweet path length. Our results found an average of 5, while other research reported metrics between 2 and 4.8. This finding suggests a greater spread of suicidal ideation content than that reported in the related studies. Another point of interest with this result is that this is similar to the interaction measures reported by a very large Twitter network of over 3 million nodes (avg. shortest path 5.45), thus providing evidence of properties of large scale communication networks within a very small network and suggesting a high level of propagation of such content within the virtual community and some potential for information spread.

The *retweets graph* was composed of highly disconnected hubs (usually of small size) that propagate suicidal content between small networks via a number of users acting as bridges, demonstrating a potential for information cascade and dissemination outside the set S of authors posting suicidal intent content (with possible *contagion effect*). The relatively high shortest paths values suggest the existence of these 'weak-links'/bridges that connect together different smaller communities and, although not particularly numerous, can provide a route to propagation. While content is posted by suicidal users, retweeters are (in the majority of cases) external nodes (i.e. not posting suicidal ideation), thus allowing content dissemination outside our initial group of suicidal users. Once again, this provides evidence of a contagion effect, which has been long recognised in the suicidology field. The findings have implications for suicide prevention and especially the urgent need to develop and evaluate online interventions [36].

7. Future work

While we have identified some interesting and promising results, future research is needed in order to overcome the limitations of our analysis, conducted on a limited size set of annotated posts. In fact, even if we started from a relatively large dataset, the posts classified as containing suicidal intent did not appear to be included in large percentages (only about 10% of tweets harvested using suicide-related keywords) because of the inherent characteristics of this type of users and content. We have developed a machine classification method that is able to automatically distinguish between text containing suicidal ideation and other forms of suicidal communication, and could be used to derive a much larger dataset from social media streams for further validation and experimentation [37].

Furthermore, the analysis could be extended to more than one-hop-away neighbours (friends of friends, retweeters of the retweeters), and then to look at the characteristics of these two-and-more-hops neighbours. For example, by analysing samples of their timeline Tweets, we can investigate if, beside retweeting suicidal content, these users may have posted a similar type of content and could also be classified as 'suicidal' users (using the machine classification method in [37]). Further insights could also be derived by analysing the demographic characteristics (such as age and gender) of this type of users and their social network of friends, followers, and retweeters.

Finally, it would be also interesting to extend this study by conducting a similar analysis over a longer term, by increasing the duration of the data collection and looking at the regularity and periodicity characteristics of such content. This would allow for the investigation of the evolution of suicidal content over a longer period of time and for further reflections on the social networks of these users, perhaps including comparison with other social movements (see [35] for reference).

Acknowledgement

This research is funded by the Department of Health Policy Research Programme (Understanding the Role of Social Media in the Aftermath of Youth Suicides, Project Number 023/0165), and by the Children & Young People's Research Network as part of the research infrastructure for Wales funded by NISCHR, Welsh Government.

References

- [1] J. Pirkis, R.W. Blood, Suicide and the media, *Crisis: J. Crisis Interv. Suicide Prev.* 22 (4) (2001) 155–162.
- [2] M. Gould, P. Jamieson, D. Romer, Media contagion and suicide among the young, *Am. Behav. Sci.* 46 (9) (2003) 1269–1284.
- [3] K. Daine, K. Hawton, V. Singaravelu, A. Stewart, S. Simkin, P. Montgomery, The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people, *PLoS One* 8 (10) (2013) e77555.

- [4] L. Sloan, J. Morgan, P. Burnap, M. Williams, Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data, *PLoS One* 10 (3) (2015).
- [5] C. Haw, K. Hawton, C. Niedzwiedz, S. Platt, Suicide clusters: a review of risk factors and mechanisms, *Suicide Life-Threat. Behav.* 43 (1) (2013) 97–108.
- [6] H.-H. Won, W. Myung, G.-Y. Song, W.-H. Lee, J.-W. Kim, B.J. Carroll, D.K. Kim, Predicting national suicide numbers with social media data, *PLoS One* 8 (4) (2013) e61809.
- [7] J. Jashinsky, S.H. Burton, C.L. Hanson, J. West, C. Giraud-Carrier, M.D. Barnes, T. Argyle, Tracking suicide risk factors through Twitter in the US (2013).
- [8] P.S. Bearman, J. Moody, Suicide and friendships among american adolescents, *Am. J. Public Health* 94 (1) (2004) 89–95.
- [9] I. Kawachi, G.A. Colditz, A. Ascherio, E.B. Rimm, E. Giovannucci, M.J. Stampfer, W.C. Willett, A prospective study of social networks in relation to total mortality and cardiovascular disease in men in the usa., *J. Epidemiol. Community Health* 50 (3) (1996) 245–251.
- [10] M.S. Gould, Suicide and the media, *Ann. NY Acad. Sci.* 932 (1) (2001) 200–224.
- [11] U. Magne-Ingvar, A. Öjehagen, L. Träskman-Bendz, The social network of people who attempt suicide, *Acta Psychiatr. Scand.* 86 (2) (1992) 153–158.
- [12] I. Spasić, P. Burnap, M. Greenwood, M. Arribas-Ayllon, A naïve bayes approach to classifying topics in suicide notes, *Biomed. Inf. Insights* 5 (Suppl 1) (2012) 87.
- [13] J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, A. Leenaars, Suicide note classification using natural language processing: a content analysis, *Biomed. Inf. Insights* 2010 (3) (2010) 19.
- [14] M.A. Moreno, L.A. Jelenchick, K.G. Egan, E. Cox, H. Young, K.E. Gannon, T. Becker, Feeling bad on facebook: depression disclosures by college students on a social networking site, *Depress. Anxiety* 28 (6) (2011) 447–455.
- [15] M. De Choudhury, S. Counts, E.J. Horvitz, A. Hoff, Characterizing and predicting postpartum depression from shared facebook data, in: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, 2014, pp. 626–638.
- [16] S. Balani, M. De Choudhury, Detecting and characterizing mental health related self-disclosure in social media, in: Proceedings of CHI'15: 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, p. to appear.
- [17] H. Shawl, G.M. In defense of the internet, the relationship between internet communication and depression, loneliness, self-esteem, and perceived social support, *Cyber Psychol. Behav.* 5 (2) (2002) 157–171.
- [18] M. Merolli, K. Gray, F. Martin-Sanchez, Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances, *J. Biomed. Inf.* 46 (6) (2013) 957–969.
- [19] C. Poulin, B. Shiner, P. Thompson, L. Vepstas, Y. Young-Xu, B. Goertzel, B. Watts, L. Flashman, T. McAllister, Predicting the risk of suicide by analyzing the text of clinical notes, *PLoS One* 9 (1) (2014) e85733.
- [20] A. Abboute, Y. Boudjeriou, G. Entringer, J. Azé, S. Bringay, P. Poncelet, Mining twitter for suicide prevention, in: Proceedings of the Natural Language Processing and Information Systems, Springer, 2014, pp. 250–253.
- [21] R.C. Hsiung, A suicide in an online mental health support group: reactions of the group members, administrative responses, and recommendations, *Cyber Psychol. Behav.* 10 (4) (2007) 495–500.
- [22] D. Quercia, L. Capra, J. Crowcroft, The social world of twitter: topics, geography, and emotions., in: Proceedings of the ICWSM, 2012.
- [23] K. Lerman, R. Ghosh, Information contagion: an empirical study of the spread of news on digg and twitter social networks, in: Proceedings of the ICWSM, 10, 2010, pp. 90–97.
- [24] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, *J. Am. Soc. Inf. Sci. Technol.* 61 (12) (2010) 2544–2558.
- [25] L. Barbosa, J. Feng, Robust sentiment detection on twitter from biased and noisy data, in: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Association for Computational Linguistics, 2010, pp. 36–44.
- [26] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: Proceedings of the LREC, 2010.
- [27] P. Burnap, O.F. Rana, N. Avis, M. Williams, W. Housley, A. Edwards, J. Morgan, L. Sloan, Detecting tension in online communities with computational twitter analysis, *Technol. Forecast. Soc. Change* 95 (2013) 96–108.
- [28] S.A. Myers, A. Sharma, P. Gupta, J. Lin, Information network or social network? the structure of the twitter follow graph, in: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2014, pp. 493–498.
- [29] D. Ediger, K. Jiang, J. Riedy, D.A. Bader, C. Corley, R. Farber, W.N. Reynolds, Massive social network analysis: mining twitter for social good, in: Proceedings of the 39th International Conference on Parallel Processing (ICPP), IEEE, 2010, pp. 583–593.
- [30] J. Kunegis, Konect: the koblenz network collection, in: Proceedings of the 22nd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2013, pp. 1343–1350.
- [31] M.S. Granovetter, The strength of weak ties, *Am. J. Sociol.* (1973) 1360–1380.
- [32] D.R. Bild, Y. Liu, R.P. Dick, Z.M. Mao, D.S. Wallach, Aggregate characterization of user behavior in twitter and analysis of the retweet graph, arXiv preprint arXiv:1402.2671 (2014).
- [33] W. Webberley, S. Allen, R. Whitaker, Retweeting: a study of message-forwarding in twitter, in: Proceedings of the Workshop on Mobile and Online Social Networks (MOSN), IEEE, 2011, pp. 13–18.
- [34] V. Arnaboldi, M. Conti, A. Passarella, R. Dunbar, Dynamics of personal social relationships in online social networks: a study on twitter, in: Proceedings of the First ACM Conference on Online Social Networks, ACM, 2013, pp. 15–26.
- [35] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: densification laws, shrinking diameters and possible explanations, in: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, ACM, 2005, pp. 177–187.
- [36] N. Jacob, J. Scourfield, R. Evans, Suicide prevention via the internet: a descriptive review, *Crisis: J. Crisis Interv. Suicide Prev.* 35 (4) (2014) 261.
- [37] P. Burnap, G. Colombo, J. Scourfield, Machine classification and analysis of suicide-related communication on twitter, in: Proceedings of the 26th ACM International Conference on Hypertext and Social Media, ACM, 2015.