IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BULLYNET: UNMASKING CYBERBULLIES ON SOCIAL NETWORKS

Bleston Lawrance*, Ramesh E. R**

*(M.Sc., Department of Computer Science and Engineering, Dr. MGR Educational and Research Institute, Chennai, India

** (Faculty, Centre of Excellence in Digital Forensics, Dr. MGR Educational and Research Institute, Chennai, India

ABSTRACT

The most significant negative impact of social media is without doubt the abuse of social media. Cyberbullying appears to be even worse than traditional bullying because unlike paper records that other forms of bullying can create, records created online may remain online for a very long time, and, as it seems, there is stricter supervision. Thus, in this paper, for an efficient working mechanism, we introduce the BullyNet system that is required to recognize cyberbullies on social media pages of Twitter. Taking this bullying tendency into account, this paper aims to suggest the best method to measure a Cyberbullying Signed Network (SN). This means that the level of bullying in the tweets is determined and the tweets are properly aligned within the best possible context. Finally, the concept of centrality is included in the detection of cyberbullying in a cyberbully SN by our method and this concept shows that it is the best among the four concepts. In this work, we prove that this method helps with the identification of the culprits, where this method is adjusted for the identification of the local host of tweets with fairly good accuracy metrics and the number of messages is not as significant in this case.

Keywords: Cyberbullying, Malicious tweets, Signed network.

INTRODUCTION

Introduced several new opportunities that allowed the citizens to communicate and establish social interactions over the internet. This social media particularly has attracted a lot of attention during the past decade. Online group notification systems such as the My Space, Facebook, Twitter, flicker, and Instagram offer a different kind of interface, through which people exchange information. It is millions of people of various ages, who, waste their time on social networking sites, and hence provides a huge amount of data for the analysis as to how it can be done in case of recommenders, link suggestions to represent the scenario visually, and analyse social links. However, social media is still evolving and it continuously opens new opportunities for communication and information sharing it is also the new and potentially unsafe ground for various malicious activities, for instance, spamming, trolling, and bullying.

Cyberbullying is the use of electronic communications facilities to deliver hostile, intimidating, abusive, or hateful messages to an individual to harass, alarm, or distress them. However, the publication is written based on the experiments done in the Cyberbullying Research Centre. This is unlike normal bullying where when the act of aggression is done, be it aggression in the form of violence, it is done once or twice at the most but cyberbullying has gone on for years and the insulting messages posted are still present even to this day. These messages can be taken to the world and each of them can be a text typed by mistake, which cannot be reversed. There have also been implemented solutions in this respect but — unfortunately — there are many tools that are extremely efficient in fighting cyberbullying.

In addition, there are also report systems where users on social media can report abusive content and behaviour or misuse of the direct feature/bot to fight bullying. Towards this, we take a two-pronged strategy: first, we categorize the tweets based on tweet context to find their resemblance with Cyber-bullying & where we aim at maximizing the bullying score associated with the particular tweet other than recreating our centrality measure to identify the Cyberbullies in a Cyber-bullying SN centrality measure to detect cyberbullies from a cyberbullying SN.

REVIEW OF LITERATURE

J. Tang, C. Aggarwal, and H. Liu suggested that Recommender systems are vital, in helping users cope with the amount of information on media by offering them relevant suggestions. The rise in activities on platforms has sparked significant interest in leveraging social networks for recommendations. While most systems focus on connections in networks there is limited research on networks with both positive and negative links. The presence of links in networks poses challenges and opportunities for recommendations. We introduce an approach to utilizing signed social networks for recommendations and propose a model called RecSSN that incorporates both positive and negative links. Results from real-world data showcase the effectiveness of this framework along with experiments to explore the impact of signed networks, in RecSSN.

D. Liben-Nowell and J. Kleinberg,[4]When we look at a network at a point, in time can we predict the new connections that are likely to form among its members soon after? This is framed as the link prediction challenge. We propose methods for link prediction by assessing how "close" nodes are in a network. Studies on authorship networks indicate that insights into future connections can be derived solely from network structure and nuanced indicators of node closeness may be more effective, than straightforward metrics.

U. Brandes and D. Wagner, [5] Visone is a tool that helps with the easy viewing of social networks. Social network analysis is a way in the social sciences that uses ideas from graph theory to describe, understand, and say why social structures are the way they are. The Visone software aims to mix the looking at and the studying of social networks, and it's made for both learning and research. While it's mainly made for people studying social sciences, many of its features can also be helpful in other areas, says X. Hu, J. Tang, H. Gao, and H. Liu. In, social media was a big target for people who spam, flooding normal users with unwanted or fake stuff. These spammers make using social media for sharing and finding info hard. Unlike old spam ways, like emails and websites, social media spammers can link up easily, even if you don't agree. They join forces, pretending to be real users by getting lots of "real" friends fast. Also, the stuff you find on social media is messy and hard to sort out. Using usual ways to spot spammers on social media doesn't work well. People have looked a lot at how to catch lies in the real world through studies of how society works. So, taking a cue from what we know about how people act face to face, we're checking if understanding feelings in what people post online can help find spammers in social media. First up, we explore to see if spammers and regular folks show different emotions in their posts. Then, we come up with a new way to search for spammers online by using what we find out about these emotions.

S. Kumar,F. Spezzano, and V.S. Subrahmanian,[6] stated that online social networks such as Slashdot deliver useful information to millions of users – yet their Information is as credible as the users of such applications. Unfortunately, there are many ‘trolls’, on Slashdot who post deceitful information and affect system security. In this paper, an algorithm known as TIA (Troll Identification Algorithm) is proposed. Developed for the field of online “signed” social networks, it is aimed at categorizing users based on malicious truth. g. groups (for example, trolls on Slashdot) or the secondary, and friendly (i. e. normal honest users). While it is generally applicable to any signed social network, TIA has been evaluated for troll detection in Slashdot Zoo with the method across a broad range of parameter values. Compared to some prior algorithms, its running time is shorter and its accuracy level is notably higher than the existing one.

Natarajan Yuvaraj, [7] proposed that Recent studies have also suggested that cyberbullying is a new type of social disorder that has become an epidemic among the youths. In this paper, we propose an innovative model of textual classification of cyberbullying automation not necessitating transformation to large feature space. On the other hand, a classifier is unable to give a limited convergent solution as it suffers from the overfitting problem. For example, taking into account such limitations, we devised a text classification engine that introduces the data pre-processing step of pre-cleaning the tweets, erasing noise and other background information, extracting the selected features, and then classifying without over-fitting. The proposal of a new Deep Decision Tree classifier is made in the context of the current study, whereby a typical Decision Tree is created in a way that the hidden layers of Deep Neural Network (DNN) are used to represent the Tree node to process the input elements. The validation equally celebrates the new Deep classifier with a higher degree of accurate classification of text data.

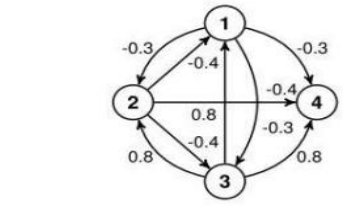
Aditya Desai, [8] it was postulated that Some Trends characteristic of background usage of the internet and related social media include the frequency of sending, receiving, and publishing negative, damaging, false, or malicious messages about another person which may be regarded as cyberbullying. Harassment in the form of cyber is also similar to threatening, slanderous, and threatening the person. Cyberbullying has had a very negative impact on people's mental health especially those of the youth as the majority of them originate from such acts. It results in low self-worth, and suicidal intentions and hence ought to be addressed. If one fails to take an appropriate and immediate measure against the increasingly popular cyberbullying phenomena, the issues of low self-esteem and unstable psychological state will affect an entire generation of youth. Some of the pipe-lined machine learning techniques that have been used in the past towards the identification of cyberbullying on social media include the following techniques. However, these models have totally and completely excluded these aspects, which could be harnessed in differentiating between bullying statements, posts, or anything and that which is not. Thus, in this paper, suggestions based on several factors that have to be considered at the time of cyberbullying identification can be proposed and determine rather important features with the help of a bidirectional deep learning model which is called BERT.

Habib Mohammed Yamani, [9] Bullying through the use of cyberspace by bullies is a real idea in every social media platform. Previous approaches to the task of cyberbullying detection have at least one of the three drawbacks below. First, they still only focus on just one SMP. Second, they address just one topic among the three identified as vital concerning cyberbullying. Third, to achieve high accuracy, they rely on the selected handcrafted features of the data. Here, we identify three main bottlenecks that can be solved by deep learning-based models. That is experiences garnered by these models in one set can be applied in another set. We performed extensive experiments using three real-world datasets: Posts: Formspring (12, 634), Twitter (15, 864), and Wikipedia (96,412). In the experiments, we get several valuable findings on the analysis of cyberbullying messages. To the best of the authors' knowledge, this is the first work to comprehensively compare and evaluate the state-of-art deep learning models and the effectiveness of transfer learning in the context of cyberbullying detection, for multiple topics and multiple SMPs.

RESEARCH METHODOLOGY

Challenges and concerns are involved if one aims to mine the social network to identify cyberbullies. First, it may be difficult to correctly understand what the particular user is implying or attempting to communicate on social media even considering the messages they post. g. where more instant and brief messages are adopted, the use of slang languages is allowed or may extend to including supplementary media such as pictures and Moving pictorial images. For instance, Twitter has shortened its' users' messages to 140 characters and could include a word or several words besides an informal word or phrase that has cultural reference meaning, emoji, and gifs. Therefore, when it comes to deciding on the performance of the classification of sentiment in a message, it becomes difficult. For this, we employ SA to decide whether the user's attitude toward other users is positive, negative, or passive. The sentiment relates to the emotion and the mood of the content while the emotion has to do with the attitude of the content. There is a wide set of libraries or tools for the classification of the text's sentiment, which includes irony, emojis, images, etc. Among all those tools, there are a few of them like VADER, Text Blob, Python NLTK, and so on.

Second, the students themselves may never care to know if they are being bullied depending on whether the bully deems it fit to become more physical, make fun of them, or just be mildly aggressive in his or her actions. Therefore, it is possible to define the intention not from the single text (message), but only taking into account the materials of the further analysis. Therefore, we have to select consecutive messages from two or more members to find the background of the user's attitude. Third, there are numerous amount of users actively connected and closely intertwined with one another on Social Internet Platforms, and because of this, it is rather challenging to distinctly identify Cyberbullies since the Social Internet Platform's structure is dynamic in nature.



Example of an SN.

Fig 1: Example of SN

To determine how one can effectively pinpoint the cyberbullies in social media, it is necessary to better understand social media as a model. The cognitive structure of relationships in social psychology is a signed graph, where a positive edge stands for the willingness of individuals to coordinate their actions for the fulfilment of individual and collective objectives, and a negative edge stands for the readiness of individuals to interfere with each other's actions for the same purpose. We make use of the signed graph and represent the Twitter social network as an SN to portray the users' behaviour, where the nodes portray the users and the directed edges portray the communication and/or relationship between the users while the weight lies within the inclusion of [-1, 1] as demonstrated in the above figure.

SSN abbreviated stands for Signed Social Network. A signed social network (SSN) is defined as a directed, weighted graph, which can be represented as G = (V, E, W), where V is the set of users or nodes, E is the set of directed edges which is a subset of V × V, the edge weights w in W are always in the range of [-1, 1]. In large networks, which most social media platforms are, it becomes quite difficult to pinpoint the cyberbullies in a network due to the dynamics and complexity involved in such a network. For instance, there are more than 500 million tweets which are posted every day on the social network of Twitter. Here, we can represent the social network in the form of a graph, and value is determined according to the malice scale of the user. This is because network analysis brings down the overly complex interaction that dictates the interaction of the users into mere nodes and edges existence.

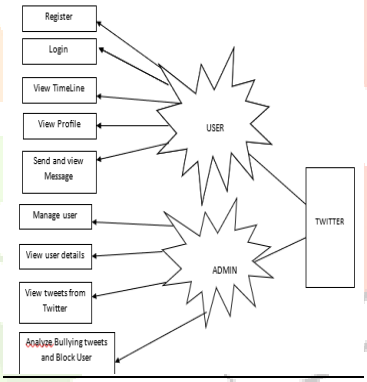


Fig 2: System Architecture

The DFD is also referred to as a bubble chart. It is a simpler graphical representation tool that may be used to describe a system based on the input data, the different processes undergone by the data, and the end output of the system.

The data flow diagram (DFD) is one of the most important modelling tools and that helps in the clarification and analysis of the process to be automated. Modelling is used to represent the specific pieces of the system under consideration. Some of these components include the system process which is the entity that performs certain tasks within the system, the data which is the content that the system process works on, an external entity which is the external entity that communicates with the system and the information flows in the system which are the paths through which information passes in the system.

It illustrates, what is received and what is broadcasted in the system, and how this data is transformed as it progresses through the various stages of DFD. Indeed, this is a graphical analysis tool that captures the flow of information as well as conversions as data is processed from an input to an output.

Another type of visual presentation is DFD which is also referred to as the bubble chart. There may exist a common behaviour such as a DFD can be developed to exhibit a system at varying levels. There is one possible division of DFD into levels that describe the system in more detail and with higher potentials of information flow.

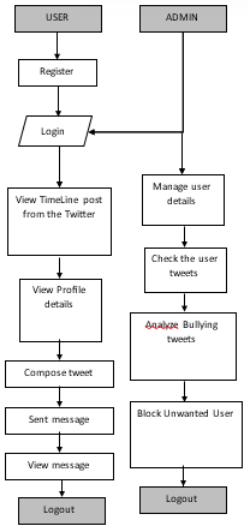


Fig 3: data flow diagram

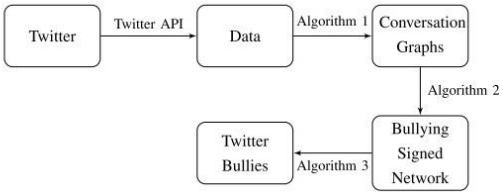


Fig 4: flowchart of bully net

From the above Twitter data, we extract conversations and the trend in the discussion of the Facebook issue considering all the variables in the model.to represent the flow of the entire conversation in a particular conversation, we create a directed weighted graph of the conversation C.{c1, c2,. . . c|C|}. In our model, the object of each ci is a set of two or more Tweets can involve two users or more, in that two or more users comment on each other or reply to one another.

Each of the account's tweets is represented by a set of tweets c = {t1,t2,. . . t|c|} for which the following condition is met. The first tweet t1 is the initiating tweet that triggers the process of the hashtag activity and then the next tweets consecutively follow the sequence. Informant of a conversation and can be assuming one of two forms:

a) DID(t1)= NULL, and dt = either MID(t1) or
Of the above fields, the occ only allows for ? NULL and either MID(t1) or &dt?.

RID(t1) is not null.

b)DID(t1)=NULL, and ∀t ⊆ T : Building our understanding using SID(t):=DID(t1).

All tweets in c satisfy the following All tweets in c satisfy the following:

SID(ti) = DID(ti+1) :In this case, integer i can be any integer that satisfies the inequality $1 \leq i \leq |c| - 1$.

It will assess the nodes and the conversions within our model. provide a list or results in an algorithm for identification of cyberbullies on the Twitter social network

The L of user-server pairs is defined as: (u|L|,s|L|)}
where ui is a user (node) and si is importance to a certain level of confidence for odds of user UI to be one of those bullies that haunt the internet.

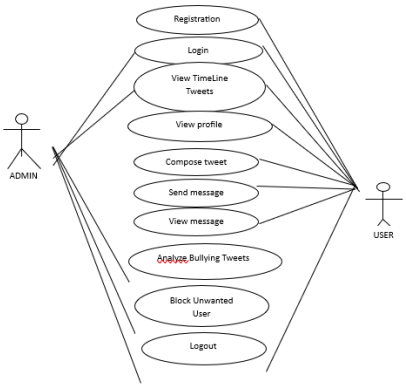


Fig 5: use case diagram

A sequence diagram is a specific interaction diagram that describes how processes interact with each other and the schedules of their actions. It is defined as the Message Sequence Chart construct. Other names associated with the sequence diagrams are event diagrams, event scenarios, and timing diagrams.

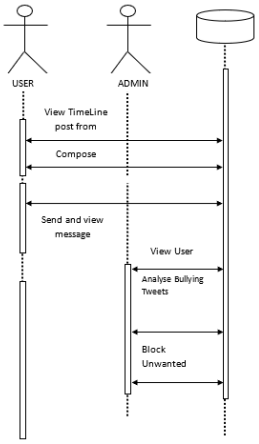


Fig 6: sequence diagram

Those activity diagrams depict the flow of activities or actions in a state by helping to display the progression of steps as well as decision, repetition, and parallel actions. The purpose that can be served by activity diagrams in Unified Modelling Language is to model the business and operational sequential flows of the components of the system. Each activity diagram generally depicts the flow of control of an activity.

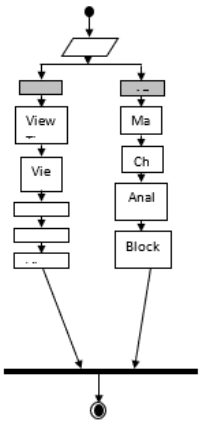


Fig 7: Activity diagram

CONCLUSION

While new-generation technology and social media networks aided improved communication platforms and interactions, societal vices such as bullying have also gained more ground. This article introduces the newly developed Bully Net approach for the identification of bully users in the Twitter social network. For a more comprehensive knowledge of how to mine SNs, we did in-depth research which helped us in establishing an SN based on bullying. We noted that when the conversation was constructed based on the context, in combination with the content, some level of management of emotions and the behaviours that are associated with bullying could be realized. Therefore, in our experiment, assessment of our proposed centrality measures in identifying the bullies from SN yields good accuracy in different cases of bully identification.

REFERENCES

[1] J. Tang, C. Aggarwal, and H. Liu, "Recommendations in signed social networks," in Proc. 25th Int. Conf. World Wide Web, Apr. 2016, pp. 31–40.https://dl.acm.org/doi/10.1145/2956185

[2] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," J. Amer. Soc. Inf. Sci. Technol., vol. 58, no. 7, pp. 1019–1031, 2007. https://cs.carleton.edu/faculty/dlibenno/papers/link-prediction/link.pdf

[3] U. Brandes and D. Wagner, "Analysis and visualization of social networks," in Graph Drawing Software. Amsterdam, The Netherlands: Elsevier, 2004, pp. 321–340. https://www.researchgate.net/publication/227172053_Statistics_for_the_Dynamic_Analysis_of_Scientometric_Data_The_evolutionof_the_sciences_in_terms_of_trajectories_and_regimes

[4] X. Hu, J. Tang, H. Gao, and H. Liu, "Social spammer detection with sentiment information," in Proc. IEEE Int. Conf. Data Mining, Dec. 2014, pp. 180–189. https://asu.elsevierpure.com/en/publications/social-spammer-detection-with-sentiment-information

[5] E. E. Buckels, P. D. Trapnell, and D. L. Paulhus, Trolls Just Want to Have Fun. Springer, 2014, pp. 67:97–102. https://www.researchgate.net/publication/260105036_Trolls_just_want_to_have_fun

[6] S. Kumar, F. Spezzano, and V. S. Subrahmanian, "Accurately detecting trolls in slashdot zoo via decluttering," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2014, pp. 188–195. https://dl.acm.org/doi/10.1145/3603399

[7] J. W. Patchin and S. Hinduja, "2016 cyberbullying data," Cyberbullying Res. Center, Tech. Rep. 2016, 2017. https://www.researchgate.net/publication/371171877_Social_Media_Platform_having_Bully_Free_Environment

[8] Cyberbullying Research Center. State Bullying Laws in America. Available: https://cyberbullying.org/bullying-laws

[9] D. Cartwright and F. Harary, "Structural balance: A generalization of Heider's theory," Psychol. Rev., vol. 63, no. 5, p. 277, Sep. 1956. https://www.researchgate.net/publication/223827415_A_balance_theory_approach_to_group_problem_solving

[10] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in Proc. 28th Int. Conf. Hum. Factors Comput. Syst. (CHI), 2010, pp. 1361–1370. https://www.researchgate.net/publication/334585766_A_Novel_Algorithm_to_Compute_Stable_Groups_in_Signed_Social_Networks