International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Identifying Cyberbullies on Social Networks

[1]PRADHYUMNA B V, [2]POORNACHANDRA S

[1]Student, Department of Master of Computer Application, East West Institute of Technology (VTU), Bangalore, Karnataka, India
[2]Assoc. Professor, Department of Master of Computer Application, East West Institute of Technology (VTU), Bangalore, Karnataka, India

Abstract

The rise of cyberbullying, which will be more terrible than conventional pestering in general given that web-based records typically persist on the Internet for a long time and are challenging to regulate, is one of the most detrimental effects of social media. We present Bully Net, a three-stage calculation for identifying cyberbullies on Twitter interpersonal organization, in this article. By offering a robust technique for constructing a cyberbullying marked network, we take use of tormenting proclivities (SN). In order to streamline the harassment score, we analyze tweets to ascertain their relationship to cyberbullying while also taking into account the context in which the tweets were written.

INTRODUCTION

THE INTERNET has opened avenues for human interaction and sociability that have never been seen before. Web-based entertainment, in particular, has exploded in popularity during the last ten years. People are associating and connecting in ways that were previously inconceivable, thanks to sites like Myspace, Facebook, Twitter, Flickr, and Instagram. People of all ages frequently engage in virtual entertainment, which supplied a plethora of data for a variety of study areas, such as recommender systems. [1], link expectations [2], perception, and interpersonal organization inquiry [3]. While the development of virtual entertainment has created a fantastic platform for communication and information sharing, it has also created a

platform for retaliatory actions like spamming. [4], trolling [5], and cyberbullying [6]. When someone utilizes technology to send harassing communications, it is called cyberbullying., abuse, or undermine an individual or a group, according to the Cyberbullying Research Centre (CRC) [7]. Unlike traditional bullying, where antagonism is a one-time, face-to-face incident, Hurtful communications that are posted online for a long time constitute cyberbullying. These signals are frequently irrevocable and might be interpreted broadly. Regulationspertaining to cyberbullying and the way things are dealt with contrast starting with one spot then onto the next. For instance, inthe United States, most of

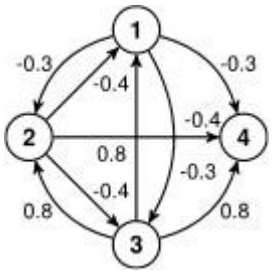


Fig. Illustration of a SN.

State harassment laws include cyberbullying, which is regarded as a criminal offence in the vast majority of them. [8]. Because of the prominence of these online entertainment locations and the anonymity that the Internet provides to the perpetrators, well- known web-based entertainment stagessuch as Twitter are

completely defenseless against cyberbullying. Despite the fact that cyberbullying is strictly prohibited by law, there are few instruments available to successfully prevent it. Clients have the option to self-report oppressive behavior and content, as well as methods to deal with harassment, on virtual entertainment stages. Twitter, for example, provides features such as account lockout. when the style of behaving becomes inadmissible or prohibiting the recordings for a certain interval to gather more knowledge and assist in the creation of efficient tools and solutions to the issue, the body of research that has been done on cyberbullying in unofficial organizations should also be expanded. We must first understand howonline enjoyment can be displayed in order to distinguish cyberbullies in virtual entertainment. The tried-and-true method of demonstrating relationship in friendly the positive edge corresponds to the grand purpose, whereas the pessimistic edge corresponds to the spiteful plan amongst individuals, according to brain science [9]. To address client behavior [10], we model the informal Twitter community as an SN, where hubs correspond to clients and coordinated edges to correspondences or maybe linkages between clients with allocated weight in the reach [11].1st Definition: A marked informal community (SSN) is a coordinated,weighted chart G = (V, E, W), where V is the arrangement of clients and E VV is the arrangement of edges in the range [1,1], with an edge weight W. Mining virtual entertainment companies for cyberbullies raises a number of issues and concerns. to decide cyberbullies forces a few difficulties and concerns. To begin with, deciphering a client's goals and meanings in virtual entertainment based solely on their communications (e.g., posts, tweets, and remarks), which are typically short, use shoptalk languages, or may include mixed media items such as photographs and sounds, is typically challenging. Twitter, Limits user communications to 140 characters, for instance, which may be a combination of text, slang, emojis, and gifs. The result is, correctly determining the assessmentprovided in a communication is difficult. We employ a feeling investigation (SA) [11], [12] to determine whether the client's attitude regarding various clients is positive, pessimistic, or nonpartisan. Second, harassment might be difficult to detect if the harasser uses methods like ridicule or aloof hostility to disguise it. In this case, In the current situation, a single instant message cannot determine the client's intent. As a result, we assemble the entire to distinguish the situation in which the client mentality exists, at least two clients should have a discussion. Third, it can be challenging to spot cyberbullies due to the size, dynamic, and complex structure of online entertainment networks. On Twitter, for example, a large number of tweets are exchanged on a regular basis on theinterpersonal organization stage. In this case, we create a diagram of the informal society and assign esteem based on the client's malignancy. as a result of the organization investigation's reduction of the complex client connection to the simple existence of hubs and edges [10], Local area location [13], hub categorization [14], and connect expectation [2] are some of the w

works that have been written about identifying harmful clients from unregistered organizations with positive

edge loads. Techniques for studying SSNs, on the other hand, are scarce [15]

II. RELATED WORKS

We'll examine at the research on cyberbullying recognition and social security numbers in this part (SSNs).

A. Cyberbullying Detection In the literature, there aren't many studies that use SNs to detect cyberbullies. The studies [6] and [17] are concerned with locating savages in an SN. Wu et al. [17] proposed a method for identifying hubs to distinguish savages without employing a PageRank computation. Kumar et al. [6] devised an iterative calculation for locating savages that includes extra cleaning chores and multiple centrality measures. The authors start their cycle with a generic SN, rather than the approach proposed in this article. A substantial amount Over the previous ten years, a significant amount of work in the area of cyberbullying identification has been completed. There are two major strategies for recognizing threats: one focuses on locating harassing texts [18]–[21], while the other seeks to locate the cyberbullies who are behind the messages. [22]-[25]. In the beginning, agonizing communications were identified using a combination of message-based inquiry and message and client characteristics. In order to categorize cyberbullies, Zhao et al. [18] proposed the message-based embeddings enhanced sack of-words (EBoW) model, which combines harassing traits, bag-of-words, and dormant semantic aspects to provide a final picture. Rather of evaluating whether a message was tormenting, Xu et al. [21] used text-based data to identify feelings in harassing follows. For cyberbullying recognition, Singh et al. [19] presented a probabilistic sociosexual data combination. This combination combines text-based highlights, such as the thickness of bad words and grammatical feature labels, as well as informal organization features obtained from a 1.5 self-imageorganization. To distinguish cyberbullying incidents, Hossein Mardi et al. [20] used photos and text. The highlighted content and images were taken from media. meetings that included photographs and related remarks, which were then sorted into various classifiers. Cheng et al. [25] suggested a novel method for detecting cyberbullies in a multimodal environment. Kao et al. developed a model to understand cyberbullying. [26] proposed a system by concentrating on friendly job discovery. Utilizing the social facts of a meeting, words and remarks, transitory features, and companion influence the structures Cheng et al. [27], [28] suggested for locating cyberbullies.

TABLE

COMPARATIVE EVALUATION OF THE MAIN FEATURES IN RELATED APPROACHES INCLUDING OUR PROPOSED APPROACH

Approach	Detect			Attributes based on				Signed Network		Dataset			
	Cyberbullying Message	User	Other	Content	Context	User	Network	Yes	No	Twitter	YouTube	Slashdot	Instagram
Zhao *et al.* [18]	●			●					●	●			
Xu *et al.* [21]	●			●					●	●			
Hosseinmardi *et al.*, [20]	●			●					●				●
Dadvar *et al.*, [35]	●			●					●		●		
Dinakar *et al.*, [36]	●			●		●			●		●		
Squicciarini *et al.* [22]		●		●	●	●	●		●	●			
Chen *et al.* [37]		●		●					●		●		
Galán-García *et al.* [23]		●		●		●			●	●			
Chatzakou *et al.* [24]		●		●		●	●		●	●			
Mishra & Bhattacharya [34]			●				●	●				●	
Kumar *et. al* [6]			●				●	●				●	
Wu *et. al* [17]			●				●	●				●	
Ortega *et al.* [38]			●				●	●				●	
Our proposed protocol		●		●	●		●	●		●			

We discovered from the above techniques that these methodologies focus on how hostile the message's substance is based on that they identify cyberbullies but do not

Think about why the communication was unfriendly; the aforementioned papers only analyze the message's content, not the context of the entire discussion. Our approach uses the message's vocabulary to identify offensive terms, SA to ascertain the source's emotions or demeanor, and eventually an analysis of the full communication between the shipper and the recipient. These ignored elements could greatly or entirely change the outcomes of cyberbullying detection.

SSNs

This section reviews previous work on SNs [6], [10], [15], [17], [29]. Although the possibility of SNs isn't new, their application and inspection have justrecently grown. In our model, we stretchedout its use to set down hub classification. Leskovec et al. [10] have previously assessed the equilibrium and the status The status hypothesis was explored in relation to online entertainment, and a modified status hypothesis was proposed that better reflects the designs seen in SNs in virtual entertainment. Tang et al. [15], [29] performed a thorough analysis of SNs in virtual entertainment and proposed a novel classification scheme for SSN hubs. The inventors presented a method for quantitatively modelling both free and subordinate data from the connectionsusing the SN, which aggregated negative connections. In recent years, various methodologies for studying SN with both positive and negative associations have been established [30]-[33]. To adjust for negative loads on the connections, the majority of these solutions rely on basic PageRank or eigenvector centrality tweaks. Some of these acts, on the other hand, neglect how a hub's approaching edges rely on. In SNs, this refers to collaborations between approaching and active connections. Mishra and Bhattacharya who introduced inclination and merit (BAD) measures, employed this condition. A hub's worth is judged by other hubs' evaluations, whereas its dependability is measured by how well it offers reliable information about other nodes. The trial's outcome

Fig. Illustration of a tweet.

TABLE II
TWEET FEATURES

SID	DID	UID	RID	MID	Text
101	3001	UserI	UserJ	UserX,UserY	@UserX @UserY Lets meet at the central park

The BAD measures are not persuasive for spotting threats in the organization, as we can see in Section VI-D. Table I provides a rough assessment of primary education. features of related strategies, including the one we suggest.

BULLYNET ALGORITHM

In this section, we first lay out the suggested three-step menace finding calculation (BFA) in basic form before delving into the means for each stage. Based on the environment and the objects in which the tweets are discovered, our solution aims to separate the harassers from crude Twitter data. The suggested method entails three calculations based on a set of tweets T that include key Twitter information like client ID, answer ID, and so forth.

1) calculating the age of a discussion, calculating the age of a bothersome SN; and

3) Bachelor of Fine Arts. The first calculation creates a coordinated weighted conversation chart Gc by efficiently reconstructing discussions from raw

Twitter data and leveraging a more precise model of human communications. The calculation that follows generates a troublesome SN B that may be used to study consumer behavior in online entertainment. The final computation includes our suggested A&M centrality measures to identify threats from B. The interaction flow of BullyNet is depicted in Fig. 3, where the discussion chart is made for each topic using Algorithm 1 and the raw data is retrieved from Twitter using the Twitter API. The discussion diagrams are then used to create an agonizing SN.

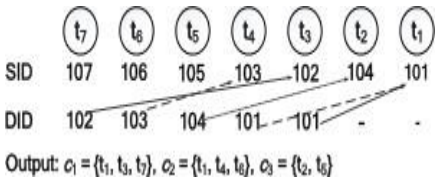
Algorithm 1 Conversation Graph Generation

Input: Set of tweets, $T = \{t_1, \ldots, t_n\}$

Output: Conversation graphs $G_c = \{g_{c_1}, \ldots, g_{c_m}\}$

1) Sort all tweets in T in reverse-chronological orc on date of creation.
2) For each tweet t_i in T, where $1 \leq i \leq |T|$:
   a) If t_i does not belong to a conversation, th a new conversation $c \in C$ and associate t_i
   b) If there is a tweet $t' \in \{t_i, t_{i+1}, \ldots, t_{|T}$ $\mathcal{DID}(t_i) = \mathcal{SID}(t')$ then associate t' wit conversations.
3) For each conversation $c_i \in C$:
   a) Construct a conversation graph $g_{c_i} \in G$ users are represented as nodes and tweets
   b) For each edge $e = (u, v)$ in g_{c_i}:
      i) Compute the sentiment of the tweet (S
      ii) Compute the cosine similarity (CS) of t with bullying bag of words (CS).
      iii) Calculate the bullying indicator I_{t_i} (w the edge as follows:
      $I_{uv} = \beta * SA + \gamma * CS$
4) Return G_c

A. Generation of a Conversation Graphis the first calculation. Algorithm 1 generates coordinated weighted discussion charts Gc for each discussion by constructing a discussion diagram age from a collection of tweets T. Byexamining the feeling behind thelanguage of a tweet and looking forrevile terms, the loads between the hubs can be determined. After that, we assign a score based on the message's articulation. For each tweet ti in T, a double pursuit DID (ti) is executed together with the SID of the additional tweets to work on the conversations. A new conversation is started if a match is found that is t and ti. In the event that a double pursuit coordinate is discovered with a tweets in ci. The diagrams are addressed as Gc =(V, E, I), where V is the arrangement of E is the arrangement of edges addressing the tweets in the discussion, and each edge is given a painful pointer esteem I as the edge weight, which is in the range of [1,+1]. When Iij = 1, it indicates that I has a negative relationship with j, and when Iij = 1, it indicates that I has a positive relationship with j. In light of SA [Valence Aware Dictionary and sentiment Reasoner (VADER)] and cosine similitudes (CSs) using a rundown ofcommonly used annoying terms, the harassing pointer for each tweet is computed as I =SA+CS. The elements of and are 0.9 and 0.1,respectively, and the inquiry hasn't completely resolved them. (see Section VI-C). 1st Model The explanation is isolated from the organization of tweets T = t1, t7 in Fig. 4. First and foremost, the tweets are ordered in ascending order, i.e., t7, t6, t1. The SID of the residual tweets is then used to search DID(t7) (t6 through t1). In t3, a match is discovered, and discussion c1 is framed. This exchange is repeated for each tweet. The c2 and c3 talks are



IV. Exploratory EVALUATION

We evaluate the presentation of the proposed computations in this section. To begin, we'll go over the data that we used in our analysis. Second, To produce ground truth, we examine the details of the execution and the processing steps. Finally, we present the preliminary results, which involve calculating the coefficients, utility, and adaptability.

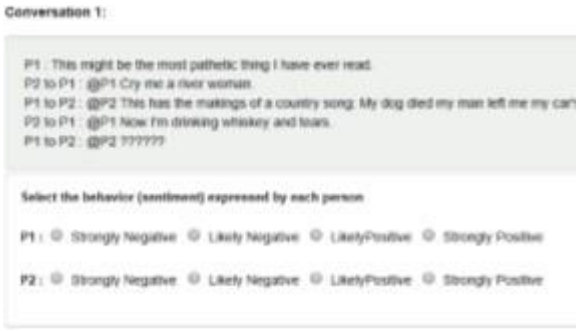


Figure 5: The Amazon Mechanical Turk study's test UI (positive: fitting way of behaving and negative: improper way of behaving).

Notices and highlights based on organizations are also supplied, including source ID and objective ID from the Twitter JSON. Only 2% of tweets were written in a language other than English. than the English language When looking at the clients, about90% of their geological area was in the United States, 6% in the United Kingdom, and the remaining 4% came from Ecuador,Japan, and China.

B. Setup and Execution

Our tests were executed on a workstation running Windows 10 64-bit, an Intel Core i7-8550U 2.00-GHz processor, with 16.0-GB RAM. Our computations were written in Java. We constructed an online review and used workers from Amazon Mechanical Turk (MTurk). Each research received 2700 overviews, each of which included ten conversations. Each overview was given tothree employees, who had to order the clients' bullying behavior in the chats according to four prepared names (emphatically certain, reasonable positive, probable negative, and firmly negative)to avoid a one-sided understanding of threats. In general, the workers looked at 27000 conversations involving 1700 people,which were extracted from the raw Twitter data using Algorithm 1. The MTurk UIallows requesters to create and distribute reviews (HITs) in a cluster, saving time by handling multiple HITs of the same type. 2700 HITs were compiled into a csv file for our assessment. For every conversational arrangement in the csv file, as seen in Fig. 7, MTurk produced a fresh HIT. The workers gave the ratings for each client related to the scheduling of meetings. Members for the evaluation came from all around the world, including the United States, Canada, Europe, and India. There was no discernible difference in the specialists' ratings. There were around 7978 unmistakably unfortunate incidents.47426 likely bad, 56704 likely certain, and 23762 firmly sure client associations. A fraction of these clients appear in a few discussions, therefore we gathered these evaluations based on the number of clients and labourers and computed using a measurement to select 569 clients as threats. Finally, the results are adjusted to create the ground truth. In a measurement, we break down and process the ground

truth, resulting in harassing and nonbullying clients. Through the verification of the suggested calculations, we assessed the exhibition measure. findings about the number of clients increasing straight from 500 to 1700, using the calculated results asthe ground truth C. Choosing the Best Values for Coefficients,, and Rememberthat Iuv = SA+CS. Suv =I uv+(IuvSvu) in Algorithm 1, and Suv =I uv+(IuvSvu) In order to calculate the coefficient and to harass pointer I and bullyingscore S, we generatedinput tweets of varying lengths and investigated various upsides of,, and, or at the very least, with 5.7 million tweets informational collection, we tested fortweets ranging from 1M, 2M, 3M, 4M, and5M for various,, and values. Following a series of experiments with various quality, we discovered that coefficient upsides of 0.6, 0.4, and 0.6 provided the best precision. Using the F1 Measure. The exactness was estimated using 0.6 and 0.4 for every 0.6 inrelation to the ground truth. The optimal qualities for the coefficients, and as well as the and values, which are set from 60 to 90, are shown in Fig. 8. and 40 to 10 for each individual We employ three alternative values ranging from 0.4 to 0.6 for each annoying pointer coefficient. According to our method, the F1 measure increases in a straight line as the coefficients increase and drop. We can also see that the F1 measure expands in every case when the value is increased, implying that the SA is more effective affects the tormenting The CS has a different marking than the CS. This is due to the fact that SA looks at the text as well as emoticons and emojis before concluding that the CS alone is harmful to the exhibition. As a result, we use both SA and Cosine. Essentially, how a person reacts to a tweet directly affects the harassment rating.

2) Precision and Recall In double classification projects, precision and recall are assessment measurements. The proportion of precision is called accuracy, and the proportion of precision is called review.

Fig. Utility concerning the quantity of clients.

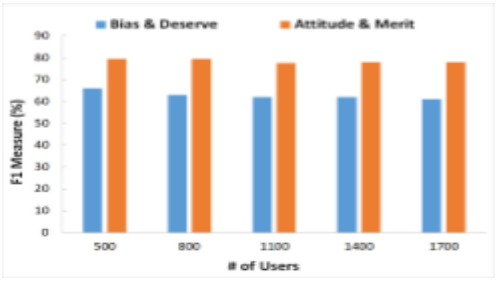


Fig. Relative assessment of A&M with BAD is the recommended centrality metric.

Then, in Section IV-C, We contrast how our proposed centrality metric A&M is presented with the research conducted by Mishra and Bhattacharya BAD. As far as precision is concerned, we look at theF1-score in relation to the number of clients generated using Algorithm We discovered that A&M has an accuracy of roughly 80% using our methods. Similarly, our centrality metrics, such as customer count, consistently beat BAD. The accuracy of BAD decreased from 65 percent to 60 percent as the number of clients increased

from 500 to 1700. A&M remains predictable despite the recommended centrality estimations. It can be used for a variety of reasons. Most notably, the propensity score of a positive hub reduces when that hub has an active edge with positive weight, and the attitude score rises when that positive hub has an active edge with positive weight. Then, when calculating the merit for a hub, the predisposition esteem is used in the range [0,1], but in A&M, the mindset esteem is used in the range [1,1]. Furthermore, when a hub has fewer amicable and approaching edges, BAD does not operate effectively. In any event, the A&M centrality has already outflanked it. With Chatzakou et al., we also examine the precision of our BullyNet calculation. Recall of Bully Net are beated in [18] and [19], separately.

TABLE

VI PERFORMANCE COMPARISONSOF DIFFERENT METHODS

	Precision	Recall	F1 Score
Chatzakou *et al.* [24]	75	53	79
Zhao *et al.* [18]	76.8	79.4	78.0
Singh *et al.* [19]	82	53	64
BullyNet	81.3	77.6	79.4

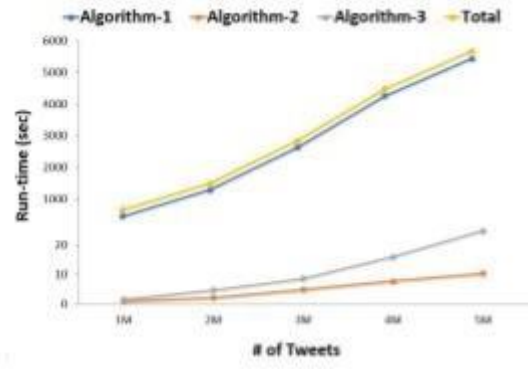


Fig. Versatility as for the quantity of tweets.

E. Flexibility We observe the run seasons of our three calculations: discussion diagrams age, harassing SN age, and menace finding with ideal qualities for coefficients, and set at 0.6, 0.9, and 0.1, respectively, to assess Bully Net's adaptability. We discovered that the Bully Net calculation takes up to 8 minutes to run an informational collection with 1 million records, and that the duration increases linearly from 1 million to 5 million records in size. Figure 11 depicts the runtime for each informational index for records sizes ranging from 1 million to five million. We also discovered that the most common estimate in our research is the age of discussion diagrams, which required some expenditure. i.e., nearly 70% of the three calculations' entire execution season. This is due to the way the discussion charts must compute SA and CS for each tweet before calculating the corresponding bullying pointer I. edge Every discussion diagram has a weight. We discovered that there is a continuous expansion throughout the runtime, as well as an expansion in various tweets. However, we discovered that the vexing SN age calculation runtime did not increase in lockstep with the increase in records, but rather remained constant. This is due to the fact that m discussion charts have k number of hubs. As a result, calculating the harassment score for each diagram takes Ok and has no bearing on the runtime as the number of tweets increases. We can see that, like the first calculation, the third computation's runtime increases in lockstep with record size. The change is attributable to a rise in the number of twitter clients, which causes the computation time for centrality indicators to increase.

RESULT

Tormenting has turned into a more common issue because of the computerized upheaval and the approach of virtual entertainment, which empowered significant leap forwards in correspondence stages and social connections. BullyNet, a clever system for recognizing menace clients on the Twitter interpersonal organization, is presented in this article. We led broad examination on digging SNs for a more profound information on the connections between clients in virtual entertainment to plan a SN in light of harassing propensities. We saw that by organizing exchanges around setting and subject, we had the option to recognize the feelings and ways of behaving that underlay harassing appropriately. In a trial examination, we tried our recommended centrality standards for perceiving menaces from SN, and we were successful in recognizing menaces in various settings.

V. CONCLUSION

Although computerized disruption and the advent of virtual entertainment enabled significant advancements in communication stages and social collaborations, harassment has become a more widespread problem. This article discusses Bully Net's innovative approach for detecting threat clients in the Twitter social media network. We conducted extensive research into SNs in order to have a deeper understanding of the connections between clients in virtual entertainment and to assemble an SN in light of annoying proclivities. We discovered that through generating talks in light of the context, as well as being satisfied, we were able to separate the feelings and behaviours that underpin tormenting. We achieved approximately 80% exactness with 81We got about 80% exactness with 81 % accuracy in our exploratory investigation of our proposed centrality metrics to detect threats from SN, and we found them to be effective. Differentiating between threats in various situations. There are a few unanswered questions that need to be investigated further. To begin, our research focuses on distinguishing emotions and behaviour in tweets from text and emoticons. In any case, examining photographs and videos would be fascinating, given that many clients use them to threaten others. Second,it fails to distinguish threats and obnoxious clients. Concocting new computations or tactics to recognise threats from aggressorswould be critical in detecting cyberbullies more effectively. Another interesting topic is to focus on the relationship between discussion diagram elements and geographic area, and what these parts are used for by the geographic dispersal of clients. Is it true that proximity increases the torturous behavior?

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.

[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.

[3] U. Brandes and D. Wagner, "Analysis and visualization of social networks," in Graph Drawing Software. Amsterdam, The Netherlands: Elsevier, 2004, pp. 321–340.

[4] X. Hu, J. Tang, H. Gao, and H. Liu, "Social spammer detection with sentiment

information," in Proc. IEEE Int. Conf. DataMining, Dec. 2014, pp. 180–189.

[5] E. E. Buckels, P. D. Trapnell, and D. L. Paulhus, Trolls Just Want to Have Fun. Springer, 2014, pp. 67:97–102.

[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.

[7] J. W. Patchin and S. Hinduja, "2016 cyberbullying data," Cyberbullying Res. Center, Tech. Rep. 2016, 2017.

[8] Cyberbullying Research Center. State Bullying Laws in America. Accessed: Jul. 1, 2020. [Online]. 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.

[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.

[11] R. Plutchik, "A general psychoevolutionary theory of emotion," in Theories of Emotion. 1980, pp. 3–33.

[12] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[13] L. Tang and H. Liu, "Community detection and mining in social media,"

[14] Synth. Lectures Data Mining Knowl. Discovery, vol. 2, no. 1, pp. 1–137, Jan. 2010.

[15] S. Bhagat, G. Cormode, and S.Muthukrishnan, "Node

[16] classification insocial networks," in Social Network Data Analytics. 2011, pp. 115–148.

[17] J. Tang, Y. Chang, C. Aggarwal, and H. Liu, "A survey of signed network mining in social media," in Proc. ACM Comput. Surv., vol. 3, 2016, pp. 42:1–42:37.

[18] J. Kunegis, J. Preusse, and F. Schwagereit, "What is the added value of negative links in online social networks?" in Proc. 22nd Int. Conf. World Wide Web (WWW), 2013, pp. 727–736.

[19] Z. Wu, C. C. Aggarwal, and J. Sun, "The troll-trust model for ranking in signed networks," in Proc. 9th ACM Int. Conf. Web Search Data Mining, Feb. 2016, pp. 447–456.

[20] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," in Proc. 17th Int. Conf. Distrib. Comput. Netw., Jan. 2016, pp. 1–6.

[21] V. K. Singh, Q. Huang, and P. K. Atrey, "Cyberbullying detection using probabilistic socio-textual information fusion," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 884–887.

[22] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Detection of cyberbullying incidents onthe Instagram social network," CoRR, vol. 1503.03909, 2015.

[23] J.-M. Xu, X. Zhu, and A. Bellmore, "Fast learning for sentiment analysis on bullying," in Proc. 1st Int. WISDOM, 2012, pp. 10:1–10:6.

[24] A. Squicciarini, S. Rajtmajer, Y. Liu, and C. Griffin, "Identification and characterization of cyberbullying dynamics in an online social network," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw.

Anal. Mining, Aug. 2015, pp. 280–285.

[25] P. Galán-García, J. G. De La Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in Twitter social network: Application to a real case of cyberbullying," Logic J. IGPL, vol. 24, no.1, pp. 42–53, 2015.

[26] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Mean birds: Detecting aggression and bullying on Twitter," in Proc. ACM Web Sci. Conf., Jun. 2017, pp.13–22.

[27] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, "XBully: Cyberbullying detection within a multi-modal context," in Proc. 12th ACM Int. Conf. Web Search Data Mining, Jan. 2019, pp. 339–347.