**INTRODUCTION**

The Internet has created never before seen opportunities for human interaction and socialization. In the past decade, social media, in particular, has had a popularity explosion. From MySpace to Face book, Twitter, Flickr, and Instagram, people are connecting and interacting in a way that was previously impossible. The widespread usage of social media across people from all ages created a vast amount of data for several research topics, including recommender systems [1], link predictions [2], visualization, and analysis of social networks [3].

While the growth of social media has created an excellent platform for communications and information sharing, it has also created a new platform for malicious activities such as spamming [4], trolling [5], and cyber bullying [6]. According to the Cyber bullying Research Center (CRC) [7], cyber bullying occurs when someone uses the technology to send messages to harass, mistreat or threaten a person or a group. Unlike traditional bullying where aggression is a short and temporary face to- face occurrence, cyber bullying contains hurtful messages which are present online for a long time. These messages can be accessed worldwide, and are often irrevocable. Laws about cyber bullying and how it is handled differ from one place to another. For example, in the United States, the majority of the states incorporate cyberbullying into their bullying laws, and cyber bullying is considered a criminal offense in most of them [8]. Popular social media platforms such as Face book and Twitter are very vulnerable to cyber bullying due to the popularity of these social media sites and the anonymity that the internet offers to the perpetrators. Although strict laws exist to punish cyber bullying, there are very less tools available to effectively combat cyber bullying. Social media platforms provide users with the option to self-report abusive behavior and content in addition to providing tools to deal with bullying. For example, Twitter has features that include locking accounts for a brief period of time or banning the accounts when the behavior becomes unacceptable. The body of work produced by the research community with regards to cyber bullying in social networks also needs to be expanded to get better insights and help develop effective tools and techniques to tackle the issue.

To identify cyber bullies in social media, we first need to understand how social media can be modeled. The common way of modeling relationship in social psychology [9] is to represent it as a signed graph with positive edge corresponds to the good intent and negative edge corresponds to malicious intent between people. Using the signed graph, we model th Twitter social network as a signed network to represent users’ behavior [10] where nodes correspond to users, directed edges correspond to communications and/or relations between the users with assigned weight in the range [-1, 1], as illustrated in Figure 1.

Definition 1: A signed social network (SSN) is a directed, weighted graph G = (V;E; W), where V is the set of users and E \_ V \_ V is the set of edges with an edge weight w 2 W in the range of [-1,1].

Mining social media networks to determine cyber bullies imposes several challenges and concerns. First, it is typically hard to accurately interpret user’s intentions and meanings in social media based merely on their messages (e.g. posts, tweets, comments) which are typically short, use slang languages, or may include multimedia contents such as pictures and videos. For example, Twitter limits its users’ messages to 140 characters, which could be a mix of text, slangs, emojis, and gifs. As a result, it is hard to determine the opinion expressed by a message correctly. For this we utilize sentiment analysis [11], [12] to determine whether the user’s attitude towards other users are positive, negative, or neutral. Second, bullying could be hard to detect if the bully chooses to disguise it through techniques such as sarcasm or passive-aggression. In this situation, a single text (message) cannot determine the user’s intention. So, we collect the entire conversation between two or more users to identify the context in which the user attitude exists. Third, the large size and dynamic and complex structure of social media networks makes it challenging to identify cyber bullies. For example, on Twitter, hundreds of millions of tweets are sent every day on the social network platform. In this case we construct the social network as a graph and assign value based on the maliciousness of the user. Because the network analysis reduce the complex relationship between the users to the simple existence of nodes and edges [10] There are several works in the literature concerning detecting malicious users from unsigned networks with positive edge weights, including community detection [13], node classification [14] and link prediction [2]. On the other hand, methods that analyze signed social networks are scarce [15].

In this paper, we study the problem of cyber bullying in social media in an attempt to answer the following research question: Can tweet contexts (conversations) help improve the detection of cyber bullying in Twitter?. Our intuition is that each tweet should be evaluated not only based on its contents, but also based on the context in which it exists. We call such a context a conversation, which is a set of tweets between two or more people exchanging information about a certain subject. Thus, our solution consists of three parts. First, for each conversation, a conversation graph is generated based on the sentiment and bullying words in the tweets. Second, we compute the bullying score for each pair of users in a conversation graph, and then combine all graphs to create an SSN called bullying signed network (B). The inclusion of negative links can bring out information that would otherwise be missed with only positive links [16]. Finally, we propose a centrality measure called attitude & merit (A&M) to detect bullying users from the signed network KB. Our main contributions are organized as follows:

1) Collected, preprocessed and labelled the Twitter dataset.

2) Proposed a novel efficient algorithm for detecting cyber bullies on Twitter. a) Built conversation. b) Constructed Bullying Signed Network. c) Proposed Attitude and Merit Centrality.

3) Experimented on 5.6 million tweets collected over 6 months. The results show that our approach can detect cyber bullies with high accuracy, while being scalable with respect to the number of tweets.