**INTRODUCTION**

ONLINE SOCIAL networks (OSNs) are frequently flooded with scathing remarks against individuals or organizations on their perceived wrongdoing. When some of these remarks pertain to objective fact about the event, a sizable proportion attempts to malign the subject by passing quick judgments based on false or partially true facts. Limited scope of fact check ability coupled with the virulent nature of OSNs often translates into ignominy or financial loss or both for the victim. Negative discourse in the form of hate speech, bullying, profanity, flaming, trolling, etc., in OSNs is well studied in the literature. On the other hand, public shaming, which is condemnation of someone who is in violation of accepted social norms to arouse feeling of guilt in him or her, has not attracted much attention from a computational perspective. Nevertheless, these events are constantly being on the rise for some years. Public shaming events have far reaching impact on virtually every aspect of victim’s life. Such events have certain distinctive characteristics that set them apart from other similar phenomena- (a) a definite single target or victim (b) an action committed by the victim perceived to be wrong (c) a cascade of condemnation from the society. In public shaming, a shamer is seldom repetitive as opposed to bullying. Hate speech and profanity are sometimes part of a shaming event but there are nuanced forms of shaming such as sarcasm and jokes, comparison of the victim with some other persons, etc., which may not contain censored content explicitly.

The enormous volume of comments which is often used to shame an almost unknown victim speaks of the viral nature of such events. For example, when Justine Sacco, a public relations person for American Internet Company tweeted “Going to Africa. Hope I don’t get AIDS. Just kidding. I’m white!”, she had just 170 followers. Soon, a barrage of criticisms started pouring in, and the incident became one of the most talked about topics on Twitter, and the Internet in general, within hours. She lost her job even before her plane landed in South Africa. Jon Ronson’s “So You’ve Been Publicly Shamed” [1] presents an account of several online public shaming victims. What is common for a diverse set of shaming events we have studied is that the victims are subjected to punishmets disproportionate to the level of crime they have apparently committed. In Table 1, we have listed the victim, year in which the event took place, action that triggered public shaming along with the triggering medium, and its immediate consequences for each studied event. ‘Trigger’ is the action or words spoken by the ‘Victim’ which initiated public shaming. ‘Medium of triggering’ is the first communication media through which general public became aware of the ‘Trigger’. The consequences for the victim, during or shortly after the event, are listed in ‘Immediate consequences’. Henceforth, the two letter abbreviations of the victim’s nme will be used to refer to the respective shaming event.

In the past, work (e.g., [2], [3], [4], [5]) on this topic has been done from the perspective of administrators who want to filter out any content perceived as malicious according to their website policy. However, none of these considers any specific victim. On the contrary, we look at the problem from the victims perspective. We consider a comment to be shaming only when it criticizes the target of the shaming event. For example, while “Justine Sacco gonna get off that international flight and cry mountain stream fresh white whine tears b” is an instance of shaming, a comment like “Just read the Justine Sacco story lol smh sucks that she got fired for a funny tweet. People so fuckin sensitive.” is not an example of shaming from the perspective of Justine Sacco (although it contains censored words) as it rebukes other people and not her.

In this work, we propose a methodology for the detection and mitigation of the ill effects of online public shaming. We make three main contributions in this work- (a) Categorization and automatic classification of shaming tweets

(b) Provide insights into shaming events and shamers

(c) Design and develop a novel application named BlockShame that can be used by a Twitter user for blocking shamers.