Beyond Cyberbullying: Self-Disclosure, Harm and Social Support on ASKfm

Zahra Ashktorab
College of Information Studies
University of Maryland, College Park
parnia@umd.edu

Eben Haber
Couchbase, Inc.
Mountain View, CA
eben@habers.us

Jennifer Golbeck
College of Information Studies
University of Maryland, College Park
jgolbeck@umd.edu

Jessica Vitak
College of Information Studies
University of Maryland, College Park
jvitak@umd.edu

ABSTRACT

ASKfm is a social media platform popular among teens and young adults where users can interact anonymously or semi-anonymously. In this paper, we identify the modes of disclosure and interaction that occur on the site, and evaluate why users are motivated to post and interact on the site, despite its reputation for facilitating cyberbullying. Through topic modeling–supplemented with manual annotation–of a large dataset of ASKfm posts, we identify and classify the rich variety of discourse posted on ASKfm, including both positive and negative forms, providing insights into the why individuals continue to engage with the site. These findings are complemented by a survey of young adults (aged 18-20) ASKfm users, which provides additional insights into users' motivations and interaction patterns. We discuss how the affordances specific to platforms like ASKfm, including anonymity and visibility, might enable users to respond to cyberbullying in novel ways, engage in positive forms of self-disclosure, and gain social support on sensitive topics. We conclude with design recommendations that would highlight the positive interactions on the website and help diminish the repurcussions of the negative interactions.

CCS CONCEPTS

•Human-centered computing → Collaborative and social computing systems and tools; Social networking sites;

KEYWORDS

ASKfm; cyberbullying; self-disclosure; topic modeling

1 INTRODUCTION

Recent years have seen a rise in the use of social media platforms that afford anonymous communication such as ASKfm and Formspring [5, 37] and mobile applications that allow anonymous sharing like YikYak and Kik [18]. While anonymous online communication has existed for decades (e.g., Usenet, anonymous chat rooms) [27], platforms like ASKfm are novel because they allow users to anonymously communicate with known recipients (i.e., semi-anonymous communication). Because anonymity has been shown to lower people's inhibitions [33], it is not surprising that these platforms have been used for cyberbullying [22, 16]. While Formspring shut down in 2015, ASKfm (which is based on Formspring's interaction model) remains quite popular among young users, suggesting that the anonymity that likely leads to problematic interactions may also enable positive outcomes for users. In this paper we examine ASKfm to better understand how users interact, as well as the impact of semi-anonymity on those interactions.We make design recommendations for semi-anonymous platforms that foster the positive interactions that lead to social support and self-disclosures. We make recommendations that would help to mitigate and diminish the repurcussions of the negative interactions on the platform.

ASKfm facilitates a variety of interactions between users with different degrees of anonymity: it allows users to follow others anonymously and to send anonymous or pseudonymous questions to specific known recipients in exchanges visible to all of the recipients' followers. ASKfm users can also non-anonymously indicate approval of an exchange (i.e., "liking" it), and give virtual "gifts." While ASKfm describes its central interactions in terms of questions (the profile post prompt is "Ask me a Question"), users' interactions are much more diverse and represent a variety of types of discourse. ASKfm should not be confused with Q&A sites that allow the posting of information-seeking questions to a broad community, soliciting answers from any member (e.g., Yahoo! Answers, Google Answers, Stack Overflow) [8] or with more general platforms like Facebook and Twitter, where interactions can include non-anonymous questions directed to an individual or broadcast to a wide audience. On ASKfm, questions are directed to a particular individual, and are often posed anonymously.

In this paper, we evaluate the following research questions:

WebSci'17, June 25–28, 2017, Troy, NY, USA
© 2017 ACM. 978-1-4503-4896-6/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3091478.3091499

RQ1: What kinds of interactions occur alongside cyberbullying discourse on ASKfm?
RQ2: How does anonymity on ASKfm shape users' disclosure and interaction behaviors on the platform?
RQ3: What kinds of design changes can make ASKfm a safer space?

We address these research questions through two studies. First, we analyze user data collected from ASKfm–the publicly available data for each user, including basic profile information as well as exchanges with other users within an individual's network–and conduct topic modeling on these exchanges, followed by coding of data to discover the diverse kinds of interactions and discourse that occur on ASKfm. We identified 11 types of discourse posted on ASKfm, including many positive modes that provide examples of why individuals continue to engage with the site. We extend these computational findings through a survey of young ASKfm users to further understand users' motivations for disclosure and interaction, especially when they have negative experiences on the site.

This study contributes new insights into the benefits and drawbacks of online anonymity, as well as how adolescents and young adults navigate an online space that can be fraught with negativity and harm. Based on our findings, we push the discussion of anonymous interactions [18] beyond the standard focus on negative and bullying messages to consider the range of positive and negative outcomes associated with site use. While we acknowledge the importance of minimizing user risks, our study highlights how and why these sites are useful to young people, i.e., by providing an outlet for interactions that may be perceived as stigmatizing in less anonymous environments.

2 RELATED WORK

Current research on semi-anonymous websites has largely focused on automatic detection of cyberbullying [20, 14] and specific exploration of cyberbullying practices[26]. By broadening the focus to consider all potential motivations for disclosure on these platforms, we can better explain users' motivations for engagement and continued use. In order to make recommendations on how to improve the platform, we must understand the context in which the cyberbullying occurs. Below, we provide an overview of the ASKfm platform and discuss the state of existing research on anonymous and semi-anonymous social media platforms.

2.1 ASKfm: Description of Platform

ASKfm's interaction model comprises of asking questions and reacting to those questions. When navigating to a user's profile, a box prompts you to "Ask @User" a question. The question can also be asked anonymously (by checking the "Ask anonymously" box) or non-anonymously. A user's profile displays the questions that they have chosen to answer. The question is only published to the user profile if the recipient chooses to answer the question. The "Friend" feed displays all questions answered by individuals that a ASKfm user follows. In recent years, ASKfm has received significant media coverage related to cyberbullying occurring on the site; for example, a Google Search of "ASKfm" reveals the headlines "10 Frightening facts about ASKfm all parents should know" and "ASKfm: A Guide for Parents and Teachers - Webwise". These headlines are a reflection of the reputation garnered by suicide incidents reported over the years that were thought to be the direct result of cyberbullying on the website. While this reputation persists, we know that all users are not collectively engaging in cyberbullying. Other types of interaction exist on the website, and this work aims to explore these types of interactions.

2.2 Discourse on Semi-Anonymous Social Media

Previous studies on semi-anonymous Q&A websites have focused on detecting and understanding the nature of cyberbullying behaviors[20, 16]. Research on semi-anonymous websites such as Formspring and ASKfm has primarily explored negative interactions, with a specific focus on cyberbullying because of the links between cyberbullying and teen suicide [22]. For example Kontostathis et al. [20] used data from Formspring to automatically detect cyberbullying content on the site. Hosseinmardi et al. [15] explored ASKfm by examining the occurrence of "negative" words and interactions on ASKfm and found that individuals with negative content on their profiles are less active and the least sociable. They also found that as positive words increase on a user's profile, the more active and engaging that user will be. Moore et al. [26] evaluated cyberbullying and anonymity on formspring and identified aggression in both online attacks and defense posts (i.e., posts that defend the victim); they further noted that anonymity correlated positively with attacks and negatively with defense posts.

2.3 Anonymity, Disinhibition, and Online Behavior

Kang et. al. contend that ephemerality is an intrinsic part of anonymous communication application [18]. However, semi-anonymous social web applications like ASKfm do not embrace the same ephemerality as other fully anonymous social web applications since posts are recorded on user profiles. Literature on anonymous media applications further reveals that anonymity empowers individuals to disclose personal information [18]. All three types of anonymity identified by Suler are present on semi-anonymous social media platforms: users can opt to be anonymous to others; others can opt to be anonymous to a particular user; and someone's language use and writing style may further anonymize them if it is not personally identifiable. Through an analysis of personal journal blogs, Hollenbaugh et al. demonstrate that those who share photos of themselves tend to participate more in self-disclosure, revealing more information about themselves. This study makes a distinction between discursive anonymity and visual anonymity, suggesting that users believe visual cues (such as photos) to be less identifying than discursive cues (like real names) [13].

2.4 Positive Outcomes of Anonymous Disclosures

Research has identified a number of positive outcomes associated with anonymous disclosures. Self-presentation is done through self-disclosure [29], revealing personal information about oneself which is compatible with the image a person is trying to project about themselves and is an important step for the development of close relationships [19]. Kang et. al. [18] observed that a high degree self-disclosure (sharing of private personal information) occurs in anonymous mobile communications like YikYak because users felt comfortable sharing private information about themselves without the risk of being judged by their network of friends. They found comfort in the anonymity and thus were able to disclose information about themselves. Likewise, numerous researchers have identified benefits to pseudonymous health forums, especially in cases of stigmatized or rare diseases, where individuals may find it difficult to find people to talk to in their offline settings [7]. More recent work has highlighted positive uses of the social media platform Reddit for highly sensitive topics like discussions of sexual abuse [1]; in addition to pseudonymity, the site offers additional features to further separate a poster from their permanent identity (e.g., temporary accounts; see [21]). Individuals who have experienced any form of past trauma are more likely to use Web-based services when they can do so anonymously [17]. Schoenebeck contends that websites like You Be Mom, a online social outlet for mothers that allows anonymous communication provides a safe forum for moms to "trespass social norms and expectations" [30].

In summary, the existing literature on anonymous mediated interactions provides a conflicting picture. On one hand, sites that facilitate anonymous interactions may encourage cyberbullying and other negative behaviors. On the other, there is significant potential for positive outcomes, especially in the form of social support, to be generated from semi-anonymous disclosures. While research has already established the benefits of anonymous platforms to narrowly focused communities like cancer forums, we will now explore the motivations for disclosure on more general question-asking sites. In the following sections, we describe findings from two studies, including topic modeling of data from more than 40,000 ASKfm users and survey data from 243 young adults who actively used the site.

3 STUDY I: DISCOURSE DISCOVERY ON ASKFM

In our first study, our goal was to discover the the kinds of interactions and discourse that occur alongside cyberbullying discourse on ASKfm (**RQ1**). The primary mode of interaction on ASKfm involves one user sending a question or message to another user. By default, these messages are anonymous, but with an affirmative action (unchecking the box immediately under the post) the initiator can make their identity visible. The recipient can then reply to the post publicly or privately. On ASKfm, users have public profiles, yet the act of "following" or friending another user occurs anonymously – i.e., a user is aware of the number of followers but does not know who is following them [2]. A user can infer who is following them by the exchanges they receive on their profile like questions and likes on questions. Each person's feed consists of exchanges which belong to individuals whom that person follows. Each person can express approval for other people's exchanges using a "like" mechanism, and each user has a page in which their "best" exchanges are viewable, i.e., those exchanges that received the greatest number of "likes" from user's network. The "Like" mechanism is exchanged non-anonymously. This enables a mix of anonymous and non-anonymous interactions; for example, we observed anonymous posts soliciting "likes" from the broader anonymous network. Those who like the question however, are identified.

3.1 Data Collection

Studies have demonstrated that websites that allow anonymous question-asking experience greater cyberbullying [20, 14]. We therefore began the data collection process by searching ASKfm using common terms that are *unambiguously* associated with cyberbullying. Our first step was to query ASKfm through Google for variations of the terms "go kill yourself" and "go die" [12, 22]. We used this as the starting point for our data collection for two reasons: (1) to capture discourse that occurred alongside cyberbullying and these terms are unambiguously malicious and (2) to explore how individuals who have instances of such behavior on their profiles or within their network used ASKfm to engage in other types of interactions. We pulled data for subsequent analysis by crawling users who interacted with the original Google search result users (through likes and questions) and then crawling their interactions iteratively (snowball sampling).

We acknowledge that our sampling choice is limited by the fact that the "like core" in our sample are users who used a variation of terms "go kill yourself", so the data explored in this study are likely biased toward more negative forms of interaction. However, our results show that despite the source of our sample, our snowballing was able to capture a wide variety of positive discourse types in addition to negative discourse types. Furthermore, our sampling choice was influenced by media attention directed towards ASKfm centered around the many individuals who have taken their lives because of comments instructing them to using terms like: "drink bleach", "go die", and "every1 will be happy if u died" [31]. We used these terms to capture the most extreme cases of cyberbullying and the discourse that occurs alongside these negatively valenced interactions. We break down the percentage of discourse types later in this paper and demonstrate the large presence of other discourse types despite the search query starting point.

We used Google search to find these posts, with the initial search yielding 19 public profiles. Before collecting user information, we contacted ASKfm and notified them of our study and data collection process.At the time of data collection (October 4, 2013), ASKfm's Terms of Service had no restrictions on users scraping or otherwise collecting user information. The interactions on ASKfm have remained the same between 2013 and the current version with the only changes occurring in color and aesthetic. AskFM does not have an Application Programming Interface (API). We expanded our search in a snowball fashion, collecting user information from all public profiles of those users who had liked exchanges on the "best" pages of the 19 original profiles. We repeated this method of data collection until we yielded over 8 million exchanges from over 40,000 users. We collected user profile information including username, user biography, user headline, the 100 most recent exchanges with time posted, author information if applicable, and the respective answer. Additionally, we collected the 25 "best" exchanges for each user, which also includes respective time posted,

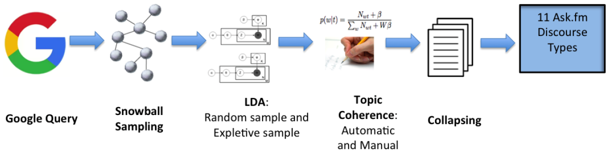

$p(w|t) = \frac{N_{wt} + \beta}{\sum_w N_{wt} + W\beta}$
11 Ask.fm Discourse Types
Google Query
Snowball Sampling
LDA: Random sample and Expletive sample
Topic Coherence: Automatic and Manual
Collapsing

Figure 1: Methodology Pipeline: Query, LDA, Topic Coherence, and Collapsing

answer, author information (if question was non-anonymous), and likers of that particular exchange.

3.2 Topic Modeling to Discover Discourse Types

Once we collected the very large corpus of posts, our next task was to discover the different types of discourse that occurred between the users. To discover different types of discourse, we approached the data using Latent Dirichlet Allocation topic modeling (LDA). Previous researchers have found topic modeling to be an efficient way to automatically discover topics, organize, and categorize large amounts of text [6, 38]. We detail how we refine the LDA process, first by using Minmo et al.'s topic coherence algorithm followed by human annotation of topics to ensure only topics that were coherent were included in the study.

While LDA is widely used for discovering topics and analyzing text, we acknowledge it's limitations. One limitation is that a user pre-defines the number of topics K. The size of K can lead to nuanced topics that overlap semantically or more general topics. Another limitation of topic modeling is that each topic is generated in the form of the most common keywords found across documents in the topic - interpretation of the meaning left to the user [3]. We addressed both limitations by 1) repeating the analysis with a wide variety of values for K, 2) using an automated measure to select those topics with the highest topic coherence values for each value of K, and 3) using multiple human annotators to validate topic coherence, label each topic, and collapse overlapping topics. The flow from data collection to the resulting topics is summarized in Figure 3.

For this analysis, we ran LDA on two samples of data. The full dataset was not analyzed because of processing limitations and a random sample would be representative of the different types of discourse. Our first sample of data consisted of 300,000 random documents from our data set. For LDA purposes, we define documents as ASKfm exchanges comprising of a question and answer combination. Our second sample was filtered for a list of expletives to contain approximately 80,000 documents. We used an expletive sample because we wanted to ensure that we captured any hint of cyberbullying in our dataset. While cyberbullying discourse may not necessarily include expletives, previous studies have found that the existence of expletives in documents are indicative of the presence of cyberbullying [20].

LDA requires, as input, a specific number K of topics to search for, and not all topics found will be coherent. Since we did not know, a priori, the number of topics covered on ASKfm, we ran LDA asking for $K = 10$, then again with $K = 20$, then with $K = 30$, all the way to $K = 100$ topics. This produced a total of 550 topics. Our next step was to automatically calculate topic coherence for each topic. Minmo et al. [25] demonstrate that calculating topic coherence is an efficient way of evaluating a topic model. They define topic coherence is the measure of co-occurrence of the top words within a topic in a document. Highly coherent topics were found within each LDA run. We wanted the data to inform our categories, so from each LDA run, we selected the most coherent 20% topics as our initial categories for qualitative coding, giving us 110 coherent topics.

We then performed a manual evaluation for each coherent topic. Our evaluators were graduate students in Computer Science who had conducted research on social networking platforms. A total of 12 annotators were recruited to annotate the coherence for each topic and rate whether a particular topic was coherent. Each topic was evaluated by three annotators. For each topic model, we presented the words in the topic and we computed the most probable documents that would fall in that topic model group. We asked evaluators whether they thought that the documents were coherent and belonged in the same category. We also asked evaluators to "in their own words" label the topic group. These labels were later used to collapse similar topic models. Cohen's Kappa inter-rater reliability score was calculated for each pair of annotators, and the total average was 0.8125 [24]. From the topic models, we selected those categories deemed coherent by all evaluators (scored higher than 3 by all three raters on a likert score), resulting in a total of 53 coherent topics.

We then analyzed the 53 user-validated coherent categories and collapsed the categories based on similarities in the labels given by annotators. For example a topic model resulting from $K = 30$ (top words: love beautiful perfect xxx amazing gorgeous aw babe girl xx thankyou sweet lovely stay he) was judged very similar to a topic model resulting from $K = 10$ (top words:love lt xxx haha xx omg thoughts hahaha amazing funny pretty aw nice cute hahah bby aha omfg). The annotators labelled both with: "complimenting a friend", "positive sentiment" and "compliments". Given the similarity in topic labels, we combined these topics into the emergent "Compliments and Positivity Discourse" category. We combined overlapping topics manually based on similarities in labels generated by human annotators. The categories represent the most coherent discourse types from the LDA sample. Combining the topic model groups based on overlapping labels resulted in 11 distinct categories (detailed below). We acknowledge that these categories do not cover *all* types of interactions, but these represent the most common modes of interaction from our sample.

Bullying/Inflammatory/Insulting Discourse:: malicious messages aimed to threaten or insult the recipient and may include in/direct threats and expletives. Responses may reciprocate inflammatory remarks.

Question: *oi mate when you go back to school am gona f****** stab you and im gona beat the shit out of you and im gona put you in hospital*

Answer: *i never new that mate come find me you c****

Compliments and Positivity Discourse:: kindness and compliments; characterized by compliments directed at the recipient.

Question: *you are so beautiful you are the nicest girl ever you have the coolest personality*

Answer: *awe c: it did make me smile .thats super nice of you to take your time and make me feel good about myself cx and umm...can i know who you are ?*

Defense of Bullied Victim Discourse:: message to cyberbullying victim in defense of previously receiving negative comment. Posters often tell victim to disregard inflammatory remarks and are sometimes include a compliment to mitigate harm.

Question: *dont listen to people who send you hate just remember that i love you and haters are gonna hate on how pretty you are its probably some fat little c*** behind a computer screen who cant say it to your face*

Answer: *phhahahhahahaha that made me laugh :)*

Like Solicitation and Rating Discourse:: asks that whoever "likes" the discourse will receive some sort of interaction on the website through "rating", "compliments", or reciprocated "likes". In the example below, each liker is promised a certain amount of reciprocated "likes" on their profile.

Question: *Likers get 5 likes and 5 questions?*

Answer: *like if you want this x*

Listing All people you follow:: asks a user to list everyone they follow on the site (via @username). This discourse type reveals "hidden" information as the site structure prevents users from seeing their followers list unless they receive a "like" interaction or are tagged in a discourse type like this one.

Question: *list of people you follow*

Answer: *@[redacted] @[redacted] @[redacted]*

Picture/Video Request:: asks for a picture/video of the recipient; sometimes coupled with a conditional that states if a user receives more than a certain amount of likes, they are deemed "pretty."

Question: *selfie?*

Answer: *[redacted].jpg*

Preference Questions:: asks about a user's preference in movies, music, pets, jewelry, etc. Answers associated with these questions tend to be straight-forward.

Question: *do you prefer gold or silver jewelry?*

Answer: *haha gold but i only have silver jewelry*

Self-Harm Discourse:: questions inquiring about someone's opinion on self harm, whether they participate in it, and how they engage in self-harm (e.g., cutting, starvation).

Question: *what is your opinion on self harming? x*

Answer: *i think its a horrible thing. for someone or something make someone feel like they should use their skin as paper. some people right now feel unwanted or ugly or fat or like they just dont belong with the world because theyre being bullied*

Sexual Content Discourse:: exchanges that are sexual in nature. These questions often ask for sexual favors or preferences in sexual exchanges.

Question: *big boobs, small butt OR small boobs, big butt?*

Answer: *small boobs, big butt*

Things that Annoy you/you Hate:: questions about users' dislikes. Answers to these question vary from hatred of things like "spiders" or hatred of things that "guys do."

Question: *:something you hate?*

Answer: *actually hate everything. i hate guys who say they hate this girl and then they text every f****** day. i hate people who always try to start shit*

Thoughts and Opinions:: asks a user's attitudes toward the question asker or mutual acquaintances; responses are expected to be honest appraisals.

Question: *opinions on [redacted] [redacted]?*

Answer: *haseena:shes soooo funnny we always get the giggles ive know.her for agesss she always knows how to mke mea laugh we hv the weirdest memories! and yea she just amazing and i tell her*

3.2.1 Classifying Discourse Types. LDA topic models predict the probability that a given document belongs to a topic. To permit automatic categorization of exchanges on ASKfm, we built a Naive Bayes classifier to assign documents to the above 11 categories. The features for the classifier were the key words identified by each topic model. For our training set, we manually annotated a sample of 1100 documents according to our derived discourse types. The collapsing of topics for each discourse type generated key words for our feature engineering process. Since a document in each of our topic models consisted of a question-answer combination (the way discourse types are presented on ASKfm and other self-anonymous social websites), our features checked for the existence of the keywords in the question-answer combination. The performance of this classifier can be seen in Table 1. The performance of our classifer is encouraging, suggesting that is it possible to perform reasonably accurate automatic detection of different discourse categories. The performance is not suprising since the LDA topics are based on keyword frequency and a predictive model based on the same keywords should be accurate.

3.2.2 Anonymity and Discourse Types. While messages sent on ASKfm are anonymous by default, senders may choose to make a message non-anonymous. To understand whether this choice was related to the type of discourse, we measured the fraction of each

Table 1: Naive Bayes Classification Results

Discourse Type	Precision	Recall	F-Measure
Complimenting/Positivity	0.987	0.865	0.922
Bullying/Inflammatory	0.798	0.753	0.775
Picture or Video Request	0.888	0.978	0.93
Preference Question	0.989	1	0.994
Like Solicitation and Rating	0.89	0.91	0.9
Thoughts/Opinions	0.971	0.382	0.548
Defense Discourse	1	0.933	0.965
Sexual Content	0.883	0.933	0.907
List of Followers	0.953	0.91	0.931
Things that you Hate	0.944	0.955	0.95
Self Harm	0.888	0.807	0.845
None of the Above	0.53	0.865	0.658
Weighted Avg.	0.886	0.858	0.856

category for which the messages were anonymous. The results are shown in Figure 2. The majority of the exchanges were anonymous. While bullying may seem to be a natural byproduct of anonymity on ASKfm, other more positive discourse is also associated with anonymity. It is worth noting that the most anonymous categories include both healthy/fun things, such as "like" solicitation and picture requests, as well as bullying. It wasn't surprising that the positive discourse types such as *Compliments and Positivity Discourse* and *Defense Discourse* were more often less anonymous than their negative counterparts like Bullying Discourse.

3.3 Limitations

We acknowledge that our sampling choice is limited by the fact that the core of our sample are users who used a variation of terms "go kill yourself" and thus captures a facet of the ASKfm usership. Our results show that despite the core of our sample, our snowballing was able to capture a wide variety of positive discourse types in addition to negative discourse types. We discovered that despite the negativity associated with ASKfm [20], the existence of the other discourse types we discovered in this study shed light on the unique affordances the semi-anonymous social media platforms offer users who are seeking social support or self-disclosing information on the website. We acknowledge that these aren't the only discourse types that occur on the website. To unpack disclosure and interaction behaviors on askFM, we describe Study II below.

4 STUDY II: ASKFM USE MOTIVATIONS

To understand users' disclosure practices and interaction behaviors on askFM ASKfm (**RQ2**) and ways the site could be improved (**RQ3**), we conducted a survey in January 2015 of young adults (ages 18-20) who identified as active site users. In our survey, we asked participants about their personal experiences with bullying and cyberbullying, their question-asking practices, as well as demographic information and measures of personality [11] and self-esteem [28]. After receiving IRB approval, we first pre-tested the items with 50 Mechanical Turkers, then opened the HIT to include up to 250 responses. In total, we received 243 usable cases for analysis. We added our age restrictions to the Mechanical Turk HIT and participants selected a box confirming they were 18-21 years old. Though this is not foolproof, it is the most we could do given Mechanical Turk ToS restrictions.

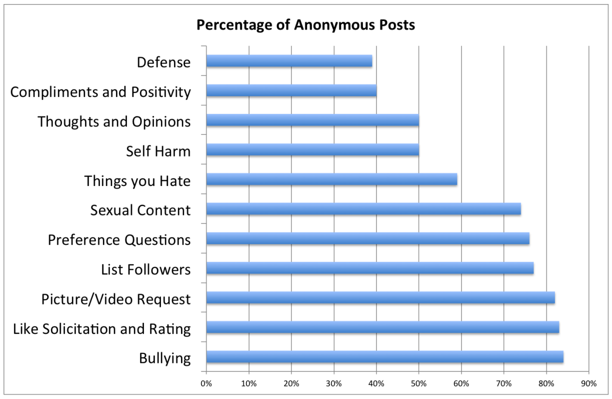


Figure 2: Percentage of Anonymous Posts for Each Mode of Discourse

4.1 Participant Demographics

In the full dataset, 35% of participants were female, and the average age was 19.6 (*SD*=.82). Two-thirds of participants were American, with the remaining participants representing 17 nations. The majority were enrolled in school full-time (60%) or part-time (15%) and lived at home with their family (57%). Participants reported spending just under five hours online per day (median=4.5 hours; *SD*=3 hours, 15 minutes), and said they used seven social media platforms on average (median = 6, *SD* = 3.76) from a list of 16 options.

4.2 Experiences with Bullying and Cyberbullying

Because popular media accounts have highlighted the prevalence of cyberbullying on ASKfm, we asked participants about their experiences with cyberbullying. Nearly three-quarters (73%) of participants reported that they had been bullied offline at some point, while 49.6% said they had been victims of cyberbullying. Conversely, when asked if they had ever participated in seven bullying activities (e.g., teasing, spreading rumors, name calling, threatening), 91.4% of participants said they had participated in at least one activity and 24.5% said they had engaged in all seven at some point. Females reported being victims of bullying (*M*=2.58, *SD*=1.12 vs. *M*=2.21, *SD*=1.09), *t*(241)=-2.52, *p*=.012, and cyberbullying (*M*=2.10, *SD*=1.20 vs. *M*=1.79, *SD*=1.02), *t*(242)=-2.15, *p*=.03, more often than men, but there were no gender differences in engaging in bullying activities. Females were also significantly more likely to experience verbal bullying than males (*M*=3.77, *SD*=1.02 vs. *M*=3.24, *SD*=1.03), *t*(172)=-3.35, *p*<.001, but no differences were reported in experiences with physical bullying.

4.3 ASKfm Interactions

The aim of our survey was to understand how ASKfm users communicated through the site, given specific affordances like anonymity and high visibility of content. We asked participants how often they interacted with strangers on the site, finding that just 7.7% of respondents reported they never interact with strangers, while more than 60% said they interact with strangers with some regularity. Given the ability to post questions anonymously on ASKfm, we asked participants if they had ever asked themselves a question anonymously, and 21% said they had. When asked about the reasons behind this practice,the most notable responses were that they did it to increase activity on their profile (71.4%), to make identity disclosures they wanted others to see (67.8%), and to cheer themselves up (54.4%). We also asked users how much they agreed with the following statement: *Posting anonymous questions on my page makes me feel better about myself*; 51.8% agreed or strongly agreed, suggesting that the simple act of posting and viewing content on their profile page–even when the content is self-written–can positively impact well-being.

We asked respondents why they think people decide to answer questions that are mean and hurtful, ultimately publishing the mean and hurtful comment to a wider audience; 54.8% said they posted the malicious comments they received because they wanted people in their network to comment on the malicious post to show support and defend them against the poster, while 52.5% said they are angry or upset and want to say that the comment is not true or they want to look like they don't care.

Finally, we asked participants if other users had ever posted malicious comments to their page. Approximately half (48.7%) said they had received negative comments about their appearance (weight, looks), 35.4% had received negative comments about their sexuality, 33.6% had received insulting comments about personal relationships, 21.4% had received "threatening comments," and 45.1% had received comments that made them feel excluded.

5 DISCUSSION

5.1 Emergent Behaviors on ASKfm

The discourse types we discovered suggest that there are other interactions and behaviors on ASKfm that occur beyond cyberbullying – interactions that are afforded by the same designs that lead to cyberbullying (anonymity for example). We found that users (1) engage in self-disclosure practices and (2) seek social support. We observe these practices across many of the discourse types we discovered.

5.1.1 Self-Disclosure on Semi-Anonymous Q&A Websites. On ASKfm, we observed very revealing acts of self-disclosure as part of *Self Harm* and *Thoughts and Opinions* discourse. *Thoughts and Opinions* discourse allows users to openly state their opinions about people mutually known by the questioner and recipient know, i.e "thoughts on Sarah?". In our survey, one highly cited reason for anonymous self-questioning was to share information that a user wanted others to see (57.1%), suggesting that anonymous self-questioning lowers the barrier to disclosing sensitive information. Self-disclosing on a semi-anonymous social media platform can be cathartic and comforting. Disclosures can occur as part of a question (usually anonymous), or in the response. What these sites offer is the opportunity to discuss something without explicitly bringing up a topic. A user might anonymously self-question, to give themselves an excuse to respond publicly, or they might ask others anonymously to recruiting those others to join in. For example, a user who anonymously posts a "suicide-list" list question can find out if anyone else in the network is experiencing the same things without identifying himself/herself.

Disclosures need not be sensitive, as we saw with *Things you Hate* and *Preference Questions. Things you hate* questions typically asked a person "What are some things that annoy you?" and a user would respond about the things that the particular person disliked. *Preference Questions* were diverse in the subject matter. They ranged from but were not limited to favorite foods, bands, mode of entertainment, and mode of communication. Questions and answers in *Things you Hate* and *Preference Questions* were innocuous invitations for users to self-disclose.

5.1.2 Social Support on Semi-Anonymous Q&A Websites. Our results demonstrate that social-support seeking behaviors are common on ASKfm. Our conceptualization of "social support" is in line with other computer-mediated communication research focused on social media and resource provision (eg, Ellison et al. found that liking behaviors were linked to bridging social capital perceptions) [9]. Support from social media can manifest through low-cost interactions (e.g., PDAs) depending on a sitefis affordances. In this section we discuss how users seek social support through: (1) self-questioning anonymously, (2) choosing to publish cyberbullying content to one's profile and (3) Like-Solicitation exchanges. The *Self-Harm Discourse* category included people seeking social support on taboo subjects like self-harm, self-injury and depression. In previous studies, users reached out for social support completely anonymously on social media platforms like YikYak or anonymous message boards [18]. On ASKfm, while a topic is brought up anonymously through the question-asking format, users can identify that they need help, or support those who need help, through a low-cost interaction of liking the post.

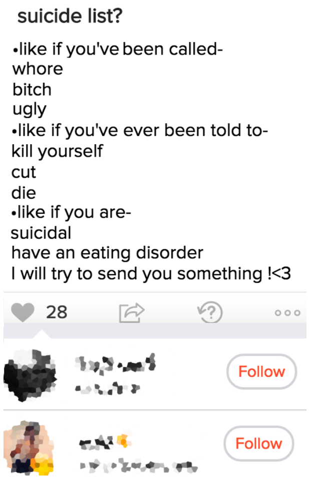
A common example of *Self-Harm Discourse* are Suicide List Questions, where the questioner asks any readers who have considered suicide to identify themselves publicly (by liking the post) in order to receive support. The poster and commenter(s) are working together to identify themselves by liking the question if they have ever been hurt in any of the ways specified by the answerer of the question. The answerer then promises to send *Compliments and Positivity* to the victims of these kinds of hate. A "Suicide List" can also qualify as a form of like bartering. Users who "like" the question are bartering for *Compliments and Positivity.* This kind of discourse demonstrates how users seek and receive social support on taboo topics.

A study looking at self-injurious behavior on message boards observed that these message boards provide essential social support, but also normalize such behaviors [36]. It is not surprising to discover taboo discourse like *Self-Harm Discourse* on a semi-anonymous social media platform. What is unique however is the transition from anonymous questioning to non-anonymous social support permitted by ASKfm–users can ask for help anonymously, but replies in support are shown visibly.

In our survey, 47.4% of self-questioners reported that they had asked themselves questions anonymously to respond to negative or harassing posts on their page, suggesting that some ASKfm users publish negative "questions" and respond by posting an anonymous "question" to their own pages as a form of support to imply they have a supportive network. Users may also be asking themselves questions in order to defend themselves when they experience cyberbullying or other negative comments. Sometimes requests for social support were more explicit (and less serious). *Like Solicitation* discourse involved users exchanging "likes" or "ratings". For example, an anonymous user posted, "likes for anyone who likes this post", and the recipient replied, "sure". In this exchange, the recipient would then have to return the "likes" for all the friends who "liked" that conversation.

5.2 ASKfm Specific Affordances

The discourse types we discovered along with our survey responses demonstrate that ASKfm affords at least three specific types of interactions: (1) Anonymous Self-Questioning Practices, (2) Transitioning from Anonymity to Visibility, and (3) Built-in Filtering of Content. We discuss each of these in detail below.


suicide list?

•like if you've been called-
whore
bitch
ugly
•like if you've ever been told to-
kill yourself
cut
die
•like if you are-
suicidal
have an eating disorder
I will try to send you something !<3

Figure 3: Example of a "Suicide List" question-answer pair on ASKfm. The question is asked anonymously. The 28 likers however are identified users. A preview of those who "liked" the question-answer pair appears at the bottom of the image. *Content has been changed to both reflect the reality of the content of these posts and to protect the identity of people involved in this post.*

5.2.1 Self-Directed Anonymous Questions. ASKfm's interaction model permits anonymous, self-directed questioning. The anonymous feature allows users to interact with themselves anonymously making it appear that they have more social support than what actually exists. This interaction is not much different than someone guilefully sending themselves flowers or gifts to their workplace anonymously to indicate to others more social approval than really exists. In our survey, 21% of the respondents said that they had asked themselves a question anonymously on their profiles, and their justifications for doing so included making identity disclosures that they wanted others to see, increasing activity on their profile, and feeling better about themselves.

While research has found that users make identified disclosures on sites like Facebook as part of their "identity work" [35], guiding self-presentation anonymously may provide users with a greater perception of control. The potential for positive effects of self-posting on perceptions of well-being is also worth further investigation; for example, researchers have found that text-based interactions have a greater positive effect on well-being than face-to-face interactions [10] while public disclosures on social media serve a self-affirming purpose by satisfying needs for self-worth [34].

5.2.2 Transition from Anonymity to Visibility. The mix of anonymous and non-anonymous interactions on ASKfm provides the ability to transition from anonymity to visibility: while "questions" can be asked either anonymously or non-anonymously, "likes" are *always* visible. Users can broach a sensitive topic anonymously, safely, and only reveal themselves through a "like" if enough of their social circle does likewise. At first glance, the use cases for such a transition might seem trivial. However, after further examination of the discourse types like *Like Solicitation*, *Compliments and Positivity*, and *Self Harm*, this transition can be an important strategy garnering social support. For example, research has found that people with lower self-esteem consider Facebook a good place to disclose information; however, they also post more negative posts, which receive fewer "likes." In turn, these users are less likely to obtain social resources from the site [9].

5.2.3 Built-In Filtering. ASKfm differs from other more conventional social media platforms because of it's implicit built-in filtering mechanism. When a user navigates to another user's profile to "ask a question", the "question" is not automatically published on the recipient's profile. Instead, the recipient receives the question in a private inbox and then can *choose* whether to respond to the message. If a user declines answering a bullying question, only the bully and the recipient know about the question. Our results reveal that users sometimes decide to answer questions even if they are hurtful or embarrassing, publicizing the hurtful question by answering it.

ASKfm's filtering allows users to reject cyberbullying and other malicious content and prevent it from becoming replicable, permanent and searchable. Users don't always reject such content, and it appears that the option to publish (or not) gives victims a degree of agency. We describe later how some users decide to publish content on their own terms, to gain social support after being bullied. This built-in filtering has implications for privacy as well. The built-in filtering allows users to consider content that may breach their privacy before it is published to their profiles. The question-asking format implicitly gives users the ability to filter *who* and *what* is being posted to their profiles.

5.3 Design Recommendations

The results of our study demonstrate that while cyberbullying is a reality of the ASKfm platform, users utilize ASKfm's affordances to transition from anonymity to visibility on taboo subjects or self-direct anonymous questions for various purposes. We can use our results to help inform better features to minimize negative interactions and possibly highlight positive interactions. We make following design recommendations:

(1) **Topic Model Filters** Our topic modeling results revealed the various words and terms that appear in cyberbullying posts. We found that top words associated with cyberbullying included a range of words like: "hate" "ugly" and "gay". The various ways these words can be interpreted based on their respective contexts demonstrate that the degree to which words can hurt depends on many factors including the context in which the word was used. Furthermore, our topic modeling results demonstrate that askFM users might use expletives affectionately and using only expletives as features in a filtering algorithm may lead false positives. For example, one document classified as "Compliments and Positivity" category was, "B****[redacted] u my bff", which was captured as this category correctly despite the fact that it included an expletive. Based on our results, we suggest that topic modeling be used to determine categories of discourse that users can then choose to filter. This approach to filtering is an alternative approach

to previous automated methods of filtering that might not take into consideration context.

(2) **Increasing Accountability by Identifying Followers:** While ASKfm users are aware of *how many* people are following them, they do not know *who* is following them on the platform. While social media users generally do a poor job at estimating the audience of their posts on more popular social media sites like Facebook [23], it more difficult to estimate who views one's content since users do not know who is "following" their posts. One of the discourse types that emerged in our first study was *Listing People You Follow*, precisely because people do not know who is "following" them. By making audiences visible to users, users will have a better understanding of who sees their content and they will be held more accountable for the content that they post or send to one another. Studies show that perception of audiences influence content production and self-presentation practices [4]. Social media platforms that allow visible subscriptions and unsubscriptions give some degree of feedback about quality of content to users, but since ASKfm's subscriptions are unknown to the user, it is more difficult to understand how people react to the content beyond other cues like reshares [32] and likes. Giving this minimal level of audience transparency to users would increase accountability on ASKfm.

6 CONCLUSION

In this work, we discovered interactions occur alongside cyberbullying discourse on ASKfm, offered reasons why people use ASKfm despite the site's propensity for cyberbullying, and concluded with making design changes that can make ASKfm a safer space. We approached these questions through a data-driven approach coupled with a survey of active ASKfm users. We used topic modeling, and manual coding to derive discourse types on ASKfm and conducted a survey and about ASKfm user. We present 11 discourse types on ASKfm that occur alongside cyberbullying discourse. We suggest that users engage in these discourse types to self-disclose and seek social support and use the affordances of ASKfm to engage in such behaviors using affordances on the platform like self-directed questions and transition from anonymity to visibility by liking anonymously-asked questions. We discussed how ASKfm's semi-anonymous and non-anonymous affordances impact the types of discourse we observed, and the types of user behavior that emerged as a result. We make design recommendations that can potentially enhance user experience by decreasing the harm caused by cyberbullying, and enhancing it's social support features.

REFERENCES

[1] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3906–3918.

[2] Ask.fm. 2013. (2013). http://ask.fm/askfm/answer/14515248650

[3] Hazeline U Asuncion, Arthur U Asuncion, and Richard N Taylor. 2010. Software traceability with topic modeling. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1*. ACM, 95–104.

[4] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 21–30.

[5] Amy Binns. 2013. Facebookfis Ugly Sisters: Anonymity and Abuse on Formspring and Ask. fm. *Media Education Research Journal* (2013).

[6] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.

[7] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity.. In *ICWSM*. Citeseer.

[8] David Dearman and Khai N Truong. 2010. Why users of yahoo!: answers do not answer questions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 329–332.

[9] Nicole B Ellison, Jessica Vitak, Rebecca Gray, and Cliff Lampe. 2014. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication* 19, 4 (2014), 855–870.

[10] Amy L Gonzales. 2014. Text-based communication influences self-esteem more than face-to-face or cellphone communication. *Computers in Human Behavior* 39 (2014), 197–203.

[11] Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[12] Ryan Grenoble. 2012. Amanda Todd: Bullied Canadian Teen Commits Suicide After Prolonged Battle Online And In School. (october 2012). http://www.huffingtonpost.com/2012/10/11/amanda-todd-suicide-bullying_n_1959909.html

[13] Erin E Hollenbaugh and Marcia K Everett. 2013. The effects of anonymity on self-disclosure in blogs: An application of the online disinhibition effect. *Journal of Computer-Mediated Communication* 18, 3 (2013), 283–302.

[14] Homa Hosseinmardi, Amir Ghasemianlangroodi, Richard Han, Qin Lv, and Shivakant Mishra. 2014a. Analyzing negative user behavior in a semi-anonymous social network. *arXiv preprint arXiv:1404.3839* (2014).

[15] Homa Hosseinmardi, Richard Han, Qin Lv, Shivakant Mishra, and Amir Ghasemianlangroodi. 2014b. Towards understanding cyberbullying behavior in a semi-anonymous social network. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE, 244–252.

[16] Homa Hosseinmardi, Shaosong Li, Zhili Yang, Qin Lv, Rahat Ibn Rafiq, Richard Han, and Shivakant Mishra. 2014c. A comparison of common users across instagram and ask. fm to better understand cyberbullying. In *Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on*. IEEE, 355–362.

[17] Ruogu Kang, Stephanie Brown, and Sara Kiesler. 2013. Why do people seek anonymity on the internet?: informing policy and design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2657–2666.

[18] Ruogu Kang, Laura Dabbish, and Katherine Sutton. 2016. Strangers on Your Phone: Why People Use Anonymous Communication Applications. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 359–370.

[19] Andreas M Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 1 (2010), 59–68.

[20] April Kontostathis, Kelly Reynolds, Andy Garron, and Lynne Edwards. 2013. Detecting cyberbullying: query terms and techniques. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 195–204.

[21] Alex Leavitt. 2015. This is a Throwaway Account: Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 317–327.

[22] Tamar Lewin. 2010. Teenage insults, scrawled on web, not on walls. *The New York Times* (2010), A1.

[23] Alice E Marwick and danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.

[24] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[25] David Mimno, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 262–272.

[26] Michael J Moore, Tadashi Nakano, Akihiro Enomoto, and Tatsuya Suda. 2012. Anonymity and roles associated with aggressive posts in an online forum. *Computers in Human Behavior* 28, 3 (2012), 861–867.

[27] J Brian Rollman, Kevin Krug, and Fredrick Parente. 2000. The chat room phenomenon: Reciprocal communication in cyberspace. *CyberPsychology and Behavior* 3, 2 (2000), 161–166.

[28] Morris Rosenberg, Carmi Schooler, and Carrie Schoenbach. 1989. Self-esteem and adolescent problems: Modeling reciprocal effects. *American sociological review* (1989), 1004–1018.

[29] Hope Jensen Schau and Mary C Gilly. 2003. We are what we post? Self-presentation in personal web space. *Journal of consumer research* 30, 3 (2003), 385–404.

[30] Sarita Yardi Schoenebeck. 2013. The Secret Life of Online Moms: Anonymity and Disinhibition on YouBeMom. com.. In *ICWSM*. Citeseer.

[31] L Smith-Spark. 2013. Hanna smith suicide fuels calls for action on ask. fm cyberbullying, cnn. (2013).

[32] Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H Chi. 2010. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Social computing (socialcom), 2010 ieee second international conference on*. IEEE, 177–184.

[33] John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.

[34] Catalina L Toma and Jeffrey T Hancock. 2013. Self-affirmation underlies Facebook use. *Personality and Social Psychology Bulletin* 39, 3 (2013), 321–331.

[35] Jessica Vitak and Jinyoung Kim. 2014. You can't block people offline: examining how Facebook's affordances shape the disclosure process. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 461–474.

[36] Janis L Whitlock, Jane L Powers, and John Eckenrode. 2006. The virtual cutting edge: the internet and adolescent self-injury. *Developmental psychology* 42, 3 (2006), 407.

[37] Haejin Yun. 2006. *The creation and validation of a perceived anonymity scale based on the social information processing model and its nomological network test in an online social support community*. ProQuest.

[38] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing Twitter and traditional media using topic models. In *Advances in Information Retrieval*. Springer, 338–349.