Tracking the Yak: An Empirical Study of Yik Yak

Martin Saveski
MIT Media Lab
Cambridge, MA, USA
msaveski@mit.edu

Sophie Chou
MIT Media Lab
Cambridge, MA, USA
soph@media.mit.edu

Deb Roy
MIT Media Lab
Cambridge, MA, USA
dkroy@media.mit.edu

Abstract

To investigate the effects of anonymity on user behavior, we conduct an empirical study of the new and controversial social app, Yik Yak. First, we examine how users use the platform, analyzing patterns in posting, popularity of yaks, and vocabulary. As a comparison, we look at posting patterns on Twitter, which has similar limitations on lengths of posts, but is public and global rather than anonymous and local. Upon a sample of 2.9M posts (1.9M yaks and 1M geotagged tweets) from 20 locations across the USA, we find that interactions on Yik Yak are specific to its location limitations and reflect the schedules of its targeted demographic, college students. Second, we test two hypotheses related to anonymity and communication: (i) whether vulgarity usage is more likely to be acceptable, and (ii) whether unique topics emerge in conversations on Yik Yak. We find that posts on Yik Yak are only slightly more likely to contain vulgarities, and we do not find any significant bias in topic distributions on Yik Yak versus on Twitter; however, differences in vocabulary and most discriminative words used suggest the need for further analysis.

Introduction

While secrecy has always had its allure, in recent years, anonymous social networks have been on the rise, expanding from online forums to mobile applications. Several platforms for smartphone users, such as Secret, Whisper, and the Insider, have emerged, and work like "The Many Shades of Anonymity"—which studies content traces from Whisper—highlight some of the differences in posting behavior when compared to public networks (Correa et al. 2015).

In the past year, Yik Yak, a social media platform founded in 2013, has gained considerable attention in the press as the subject of several controversies concerning privacy and free speech (Mahler 2015). Like Whisper, Yik Yak focuses on creating local, anonymized communities. However, it targets college campuses specifically. Other past research on anonymity on the internet—such as the work by Bernstein et al. on 4chan and /b/ focuses on web forums that are accessible to the public and cover random and broad topics, often ephemeral and/or vulgar in nature (Bernstein et al. 2011). Yik Yak differs in that its users are highly targeted, through marketing and a limited posting radius, to a specific demographic. Moreover, the platform is designed with a 200-character limit and other restrictions that guide its usage to specific use cases. It is both anonymous, and less open than freeform online forums.

In this study, we seek to summarize basic user interactions on Yik Yak in addition to examining changes in language and behavior due to anonymity.

Anonymity is linked to the "Online Disinhibition Effect," which predicts increased aggressiveness and other negative social behaviors (as in the case of online "trolls"), but conversely, can be argued to lead to more openness and willingness to discuss taboo subjects (Suler 2004). In this study, we set out to answer questions about user behaviors on Yik Yak, including: What are the temporal patterns of posting? What affects the popularity or deletion (by vote) of a post? What sort of vocabulary is characteristic of posts?

We also test the following hypotheses concerning anonymity and online disinhibition effects. H1: Vulgarity usage is more likely to be acceptable on an anonymous platform (Yik Yak) vs. a public one (Twitter). H2: Unique topics (potentially taboo ones) emerge on an anonymous platform that are undiscussed on a public one.

For a baseline, we compare our analyses of Yik Yak with geotagged, public data from Twitter.

Yik Yak

Yik Yak, the focus of our study, is a Twitter-like anonymous social smartphone application. It functions similarly to an online community bulletin board. Within a 10 mile radius, users can leave posts (yaks) of up to 200 characters to the community, and interact with other community member's posts as well. Outside of the radius, yaks are read-only with the Peek feature. The community determines the popularity and persistence of yaks with upvotes and downvotes; if a post receives -5 votes, it is deleted, creating a self-selecting mechanism for filtering and censoring content. The app is designed to encourage users to enforce this mechanism: If users engage with content (upvote/downvote posts) or post yaks that get upvoted they win *Yakarma* points.

At the time of the study, Yik Yak was completely anonymized—with no persistent identities. (Currently, icons on a single comment thread link users who post more than once.) Additionally, it is localized, operating on a 10-mile

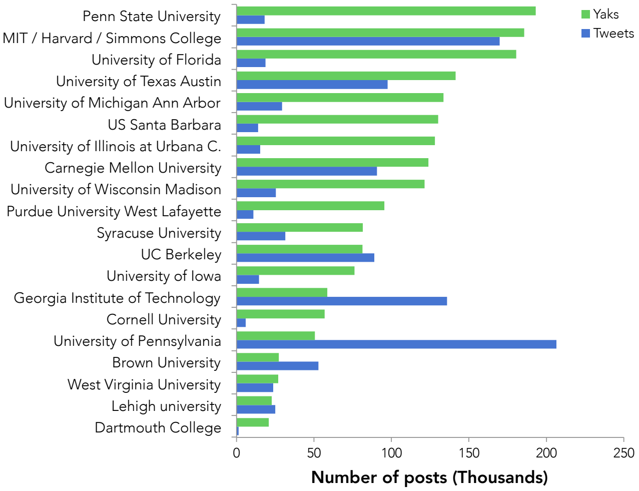

Figure 1: Locations selected for the study and the volume of yaks and tweets in each.

radius, mostly surrounding college campuses. Users can upvote, downvote, and reply to yaks. The target demographic is young, with marketing efforts heavily directed towards college campuses (Shontell 2015).

Conversely, Twitter enables users to send short 140-character limited tweets. Users on Twitter have a persistant username and identity, which can be a pseudonym or a real id, or an organization. A portion of users are "verified", or tied to their real ids. Specific to Twitter are various conventions such as @mentions, #hashtags, and the sharing of urls, all of which are not available on Yik Yak. The demographic of Twitter is less specific than the targeted audience of Yik Yak. Also, Twitter does not provide a self-selecting mechanism for filtering and censoring content, but allows users to favorite tweets and report posts that they find inappropriate.

Although, Yik Yak and Twitter are different in many aspects, using Twitter as a baseline allows us: (i) to put the results in perspective, and (ii) help future researchers understand how their existing work on Twitter—which is commonly studied—relates to this new platform.

Data Collection

We collected data from 30 hotspots over a duration of 19 days: Monday, April 4th to Friday, April 24th, 2015. For variety, we sampled from Newsweeks lists: top 10 engineering schools, top 10 women's colleges, top 10 party schools, and the Ivy League. After data collection, we narrowed the list down to 20 colleges due to sparse data in some locations, and an overlap of radiuses in others. Figure 1 shows the final set of locations and the volume of yaks and tweets in each.

For Yik Yak, we set up scrapers to run in 5-minute intervals, with a 12-hour lookback to collect replies. For Twitter, we approximated Yik Yak hotspots by looking at geotagged tweets in a 10-mile radius around the same locations. We used the GNIP Historical API to retrieve all public tweets posted in these locations. Finally, we ended up with 1.9M yaks (569K posts and 1.4M replies) and 1M tweets.

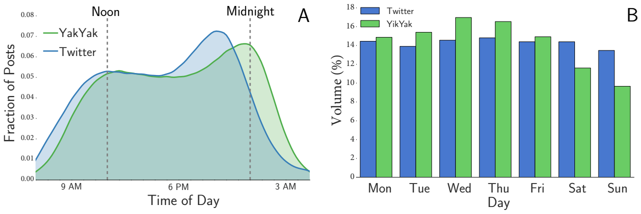

Figure 2: Daily (A) and weekly (B) activity patterns on Yak Yak (green) and Twitter (blue).

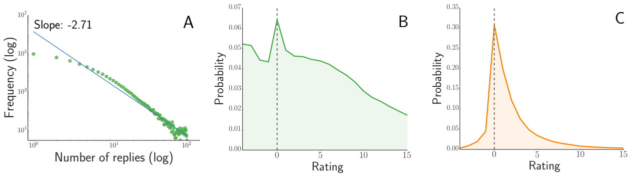

Figure 3: Yik Yak. (A) Distribution of number of replies per post. (B) Distribution of ratings (number of upvotes/downvotes) of posts. (C) Distribution of ratings of post replies.

Temporal Activity Patterns

Daily patterns. Figure 2A shows the posting activity on Yik Yak and Twitter during different times of the day. Yik Yak users start to post around 8-9am, reaching a plateau around noon. There is a steady volume of posts during the afternoon, followed by a peak around midnight. The volume decreases late at night. In contrast, Twitter users tend to post earlier in the day, starting from 7-8am. The peak of the activity is around 9pm, three hours earlier than on Yik Yak.

Weekly patterns. Figure 2B shows the posting activity during different days of the week. The volume of tweets is steady across the week, with a slight decline of 1% on Sundays. In contrast, the volume of posting activity on Yik Yak is highest on weekdays, with a peak on Wednesdays and Thursdays, and declines during weekends. This may be specific to Yik Yak's target demographic, college students, who travel home during the weekends. It is worth noting that we collected posts from only 19 days and that these patterns may vary depending on the time of year.

Yik Yak Posts and Replies

Replies. As mentioned in the previous sections, users on Yik Yik can reply on posts. We find that the distribution of number of replies per posts follows a power law (Figure 3A). In other words, most yaks have few replies and there are a few posts with a lot of replies. The platform is relatively social: every post has on average two replies. In comparison, out of all tweets we collected every third was a reply (note that we did not explicitly collect the replies of the tweets).

Ratings. Users can also rate (upvote or downvote), both posts and replies. Only a small fraction, 6.4%, of posts are not rated, meaning they have a rating of zero (Figure 3B). Most posts, 74.4% have positive ratings and only 19.2% have negative ratings. It is worth noting that we were not able to capture banned posts, i.e., posts with ratings below

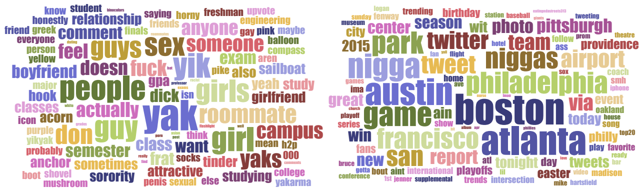

Figure 4: Words characteristic for Yik Yak (left) and Twitter (right). The size of the words is proportional to how characteristic they are to the specific platform.

-4, as they immediately disappear from the timeline. This suggests that the rating feature is heavily used, which leads to community content filtering. In contrast, about one third of all replies are not rated (Figure 3C). Those that are rated, are mostly rated positively (61.5%), and only 7.7% have negative ratings.

Content Analysis

Yaks versus tweets. Next, we analyze the differences in language use between Yik Yak and Twitter. We use Mutual Information (Hutter 2002) to find the most characteristic words for each platform. This metric selects words that are frequent, but also distinctive for the specific platform. Figure 4 shows the top 100 most characteristics words for each platform. The size of the words is proportional to their Mutual Information score, thus the more characteristic a word is to Yik Yak or Twitter, the larger it appears.

The Yik Yak word cloud is dominated by words specific to college life, such as: campus, roommate, studying, girls, guys, relationships. In contrast, the most characteristic words for Twitter are names of cities: Philadelphia, Boston, Austin, and sport related words: team, game, playoffs.

Upvoted versus downvoted yaks. In a social app with no linked identities such as Yik Yak, there is a question of whether or not community filtering mechanisms still operate to censor harmful content. Although only the user who posted the yak can delete it at will, the Yik Yak community can also remove content that is unpopular by downvoting: yaks with a cumulative sum of -5 votes are removed. By examining the most characteristic words of upvoted versus downvoted yaks, we are able to visually infer some basis of community filtering.

Figure 5 shows word clouds of the top 100 most characteristic words for upvoted and downvoted yaks, again selected using the Mutual Information criterion. The word cloud for upvoted yaks contains mostly general and slightly positive words, such as: thanks, luck, love. On the other hand, the word cloud for downvoted yaks contains mostly inappropriate and racist words.

Usage of Vulgarities

We collected a list of 355 vulgar words from `www.noswearing.com` and scanned all posts, tweets and yaks. Interestingly, we find that posts on Yik Yak are only slightly more likely to contain vulgar words than posts on Twitter: 6.29% of all yaks (posts and replies), and 5.38% of all tweets contain vulgar words. The difference is highly statistically significant ($p < 0.05$ on both χ^2 test of independence and McNemar test) due to the large sample size, but substantively very small, less then 1% or 17% relative increase.

On Yik Yak posts are more likely to have vulgar words (8.9%) than replies (5.2%). We also looked at how Yik Yak users respond to offensive posts and replies. We find that yaks that contain vulgar words are 38% more likely to be downvoted and have negative ratings: 14.9% of all yaks that contain vulgar words have negative ratings, whereas 10.8% of yaks without vulgar words have negative ratings. Furthermore, if we focus on yaks with lowest ratings (-4), they are 61% more likely to contain vulgar words.

This contradicts our first hypothesis: We find that on anonymous platforms users are only slightly more likely to use vulgar language than on public ones, and when they do it is not acceptable and leads to negative feedback.

Community Filtering and Rewards. These results contradict previous findings on anonymous online platforms such as: 4chan—where inappropriate language is considered acceptable (Bernstein et al. 2011), and Whisper—where 18% of all posts are deleted by moderators (Wang et al. 2014).

We posit that there are two key characteristics of the Yik Yak platform that lead to this difference in behavior: (i) a community filtering mechanism, and (ii) a reward system that enforces this mechanism and biases it in a positive way.

The empirical evidence that vulgar posts are more likely to be downvoted and eventually banned suggests that the community filtering mechanism is at work. We believe that this mechanism is fueled by the *Yakarma*: a user score on the Yik Yak app. Users win *Yakarma* points if they post or reply (+2 points), upvote or downvote other posts (+1 point), or if their posts get a reply (+1 point) or an upvote (+1 point), but lose *Yakarma* points when their posts are downvoted (-2 points). This encourages users not only to engage with content and enforce the community filtering mechanism, but also to post content that is appropriate and leads to engagement and upvotes.

The combination of community filtering and rewards creates an environment where positive social norms emerge. It fosters the positive aspects of the online disinhibition effect and sanctions the negative ones. This is in contrast to other anonymous platforms, such as 4chan, where vulgar language is the social norm and part of the user group identity.

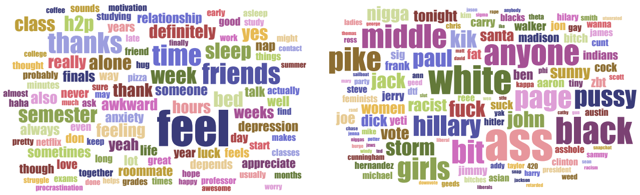

Figure 5: Top 100 most characteristic words for upvoted (left) and downvoted (right) yaks. The size words is proportional to how strongly associated they are to the class.

Topic Modeling

One of the key hypotheses regarding anonymous social platforms is the question of whether they encourage significantly different topics of discussion, either to negative (deviant) effect, or positive (encouraging) conversation. To test this assumption, we ran a topic model on the combined corpus of both tweets and yaks. Then, we assigned each document (either a tweet or a yak) to the derived topics and checked whether there are any topics that are dominated by only one platform. Using a Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) model with 200 topics, we did not see a significant amount of topics that belonged heavily to one platform and not the other. We also experimented with different number of topics and we found similar results. Figure 6 shows the sorted ratios of documents belonging to topics 1-200 in either platform.

This refutes our second hypothesis: We find that, on average, topics tend to be evenly split across both platforms and that there are no topics that are unique to Yik Yak.

Although this counters expectations of the effects of anonymity, we offer three possible explanations to this finding. (i) The bulk of both yaks and geotagged tweets center around local events and transactions. As seen in Figure 4, the majority of characteristic words for tweets are referring to cities or event venues. Although characteristic words for yaks do not include mentions of locations by name, the platform is designed specifically for local interactions. (ii) Twitter does not impose real name policy for usernames, and it may be that users use anonymous accounts (pseudonyms) to post about more sensitive topics. Previous studies show that 25.9% of all accounts are fully or partially anonymous (Peddinti, Ross, and Cappos 2014). (iii) Geotagged tweets might pose additional privacy threats to users (Mao, Shuai, and Kapadia 2011); thus users who share their location on Twitter might be less concerned about their privacy and feel comfortable to discuss sensitive topics publicly.

Discussion and Conclusion

Key findings. In our study of Yik Yak, we set out to test two main hypotheses related to anonymity and behavior. We find that vulgarity usage increases only slightly in secrecy, and furthermore, posts containing offensive language are more likely to be downvoted. This suggests that self and community censoring exists even in anonymity, and that communities on Yik Yak create a self-regulating mechanism through which vulgar comments are given negative feedback. In our analysis of content, we do not find any topics that are specific to Yik Yak. However, we see posting patterns and vocabulary that is indicative of the college-centric user base.

Limitations. Our study suffers from three main limitations. First, we are missing the content of banned yaks as they are removed immediately after they reach a rating of -5. They may contain unique topics or allow us to further investigate the community censoring hypothesis. Second, we considered only geotagged tweets, which comprise only a small fraction of all tweets and may not be a representative sample of Twitter. Notably, geotagged tweets may represent specific usages, such as declaring traveling, attendance at an event, etc. Finally, on both platforms, we collected data from a short period of time, 19 days. Patterns of behavior may also vary depending on the time of year.

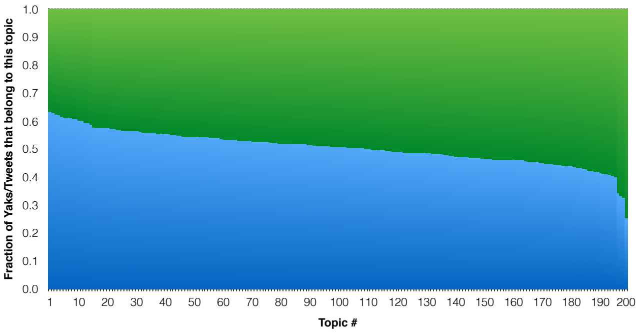

Figure 6: LDA topic modeling with 200 topics. The colors show the proportion of tweets (blue) and yaks (green) that belong to each topic. Topics are evenly split across both platforms and there are no topics that are unique to Yik Yak.

Future Work. Our analysis serves as a preliminary study of Yik Yak, and opens up many avenues of future work. In this study, we collected yaks from 20 different hotspots, but we aggregated the data in our analyses. In our next steps, we are interested in exploring how topics vary in each location, and whether any unique ones emerge specific to certain schools or regions. Similarly, looking for correlations between the characteristics of the schools (demographics, academic success, etc.) and usage of Yik Yak (linguistic characteristics, usage patterns) would show the promise of the platform as a social signal.

References

Bernstein, M. S.; Monroy-Hernández, A.; Harry, D.; André, P.; Panovich, K.; and Vargas, G. G. 2011. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In *ICWSM*.

Blei, D.; Ng, A.; and Jordan, M. 2003. Latent dirichlet allocation. *JMLR*.

Correa, D.; Silva, L. A.; Mondal, M.; Benevenuto, F.; and Gummadi, K. P. 2015. The many shades of anonymity: Characterizing anonymous social media content. In *ICWSM*.

Hutter, M. 2002. Distribution of mutual information. *NIPS*.

Mahler, J. 2015. New york times: Who spewed that abuse? anonymous yik yak app isnt telling. (Visited on 01/07/2016).

Mao, H.; Shuai, X.; and Kapadia, A. 2011. Loose tweets: an analysis of privacy leaks on twitter. In *WPES*.

Peddinti, S. T.; Ross, K. W.; and Cappos, J. 2014. On the internet, nobody knows you're a dog: a twitter case study of anonymity in social networks. In *COSN*.

Shontell, A. 2015. Business insider: How 2 georgia fraternity brothers created yik yak, a controversial app that became a ∼$400 million business in 365 days. (Visited on 01/07/2016).

Suler, J. 2004. The online disinhibition effect. *Cyberpsychology & behavior*.

Wang, G.; Wang, B.; Wang, T.; Nika, A.; Zheng, H.; and Zhao, B. Y. 2014. Whispers in the dark: analysis of an anonymous social network. In *IMC*.