Tracking the Yak

An Empirical Study of Behavior in Hyperlocal Anonymous Social Networks

Martin Saveski
MIT Media Lab
msaveski@mit.edu

Sophie Chou
MIT Media Lab
soph@media.mit.edu

1 Introduction

Human communication is rarely static: the language that we use and the gestures that we make are complex functions of the environment that surrounds us and the roles we embody. Social norms (the language a society speaks), in particular, prevent us from behaving on certain impulses.

Online social networks provide a unique opportunity to empirically study behavior and language use in different settings. Different platforms (Facebook, Twitter, Instagram, etc.) toggle different modes of communication (character limits, photos, upvotes, downvotes) and attract different demographics, triggering varied perceptions of norms. Of particular interest is what happens when we lose our persistent identities. Is behavior, on the aggregate level, drastically altered?

2 Motivation

To study this question, we examine the interactions of users in the social app Yik Yak, over the duration of two weeks in 30 locations.

Yik Yak is a Twitter-like anonymous social network with a 200 character word limit. At the time of the study, it was completely anonymized– with no persistent identities. (Currently, icons on a single comment thread link users who post more than once.) Additionally, it is a hyperlocal network, operating on 10-mile radiuses, mostly surrounding college campuses. Users can upvote, downvote, and reply to yaks. The target demographic is young.

Twitter is an online social networking platform that enables users to send short 140-character limited tweets. A user on Twitter has a specific, persistent 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 in Yik Yak. The demographic of Twitter is less directly targeted than that of Yik Yak.

As a baseline, we compare general usage of Yik Yak with that of Twitter.

There are no existing empirical analyses of anonymous social networks, especially in comparison to another network with contrasting settings. Research on platforms like 4chan and reddit examine pseudonymity, whereas identities on Yik Yak are entirely non-persistent. A few recent papers on Yik Yak exist, but they focus on the aspect of cyberbullying. The localized characteristic of the platform has also not been studied.

3 Setup

For our study, we collected data from 30 hotspots over a duration of two weeks: Sunday, April 5th to Sunday, April 25th, 2015. For variety, we sampled from Newsweeks lists: top 10 engineering schools, top 10 womens colleges, top 10 party schools, and the Ivy League. After data collection, we narrowed the list down to 24 colleges due to sparse data in some locations, and overlap of radiuses

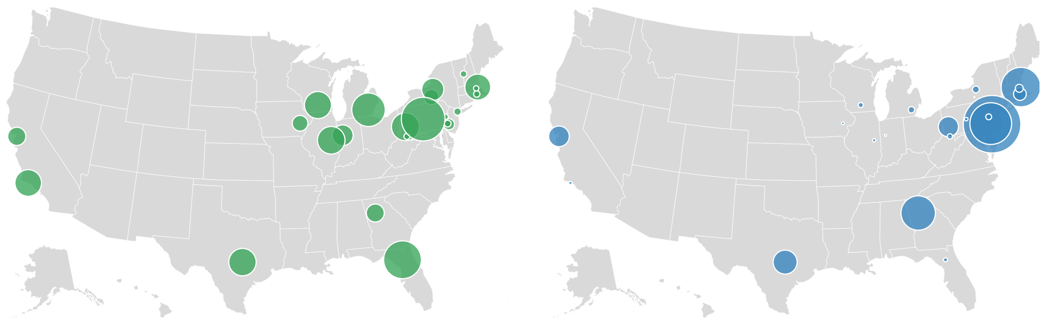

Figure 1: Hotspots for Yik Yak (green) and Twitter (blue) activity

in others. The final set of locations are shown in the maps in Figure 1 as well as the volume of yak and tweets in each location.

For Yik Yak, we set up scrapers to run in 5-minute intervals, with a 12-hour lookback to collect comments gathered. For Twitter, we approximated Yik Yak hotspots by looking at geotagged tweets in a 10-mile radius around the same locations.

4 Baseline Comparisons

To survey the general usage of the app, we measured overall behavior differences between Yik Yak and Twitter. These baseline comparisons, as well as basic statistics were covered in our class presentation and figures are included in the appendix. Briefly, we covered:

- Number of characters per post on Yik Yak and Twitter (Figure 6),
- Activity over during the different time of the day (Figure 7) and days of the week (Figure 8).
- Percentage of tweets vs. yaks containing swear words (Figure 9)
- Rating distribution of yaks (upvotes - downvotes) that contain and do not contain swear words (Figure 10).
- The fraction of banned yaks (recieving 4 downvotes) with and without swear words (Figure 11).
- Words characteristic for each platform (Figure 12).

5 Precautions

In the comparison of any two heterogenous datasets, there is a need to take precaution of potential biases. Although we survey general differences in the section above, there are still data specific concerns, including:

Population density/school size : some schools (notably Columbia University) are located in dense metropolitan areas and thus have a disproportionate amount of yaks to the school. Likewise, large state schools might have far more yaks than small colleges.

Twitter specific functionalities : there exist a number of conventions for Twitter that do not exist in Yik Yak: for example, the sharing of urls, @mentions, and hashtags.

Geotagged Tweets : geotagged tweets comprise a small (single-digit) percentage of the firehose and are often different in nature: declaring traveling, attendance at an event, etc.

Demographic differences : as stated above, the demographic of Yik Yak is far more targeted than that of Twitter and thus many differences naturally arise in language use, topic, usage patterns, etc.

All of the above have potential to be addressed in future work.

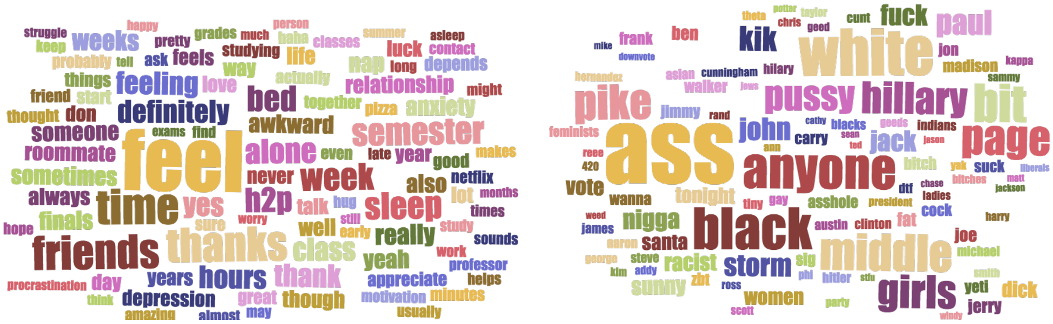

Figure 2: Top 100 most characteristic words for upvoted (left) and downvoted (right) yaks. The size words is proportional to how strongly associated they are to the class.

6 Content of up-voted vs downvoted yaks

In a social network 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 more than -4 votes are removed. By examining the top 100 words that are most predictive of upvoted yaks versus the top 100 words most predictive of community-banned yaks, we are able to visually infer some basis of community filtering. Figure 2 shows the top 100 word cloud for yaks with high number of upvotes on the left, and the top 100 word cloud for downvoted yaks on the right. The size of the term is proportional to their mutual information score, thus the more predictive a term is to upvoted/downvoted yaks, the larger it will appear.

7 Emoji usage

The use of emoji, or small pictographs, is extremely popular in Yik Yak (7.8% of all yaks contain at least one emoji) and other social messaging platforms. To gather a sense of the visual language of Yik Yak, we analyzed the usage of emojis in yaks, both in popular and unpopular yaks. In Figure 3, we show the 5 top most predictive emojis in upvoted (votes > 0) and downvoted (votes < 0) yaks. Below, we normalize the frequency of the emoji by the total number of upvoted and downvoted yaks.

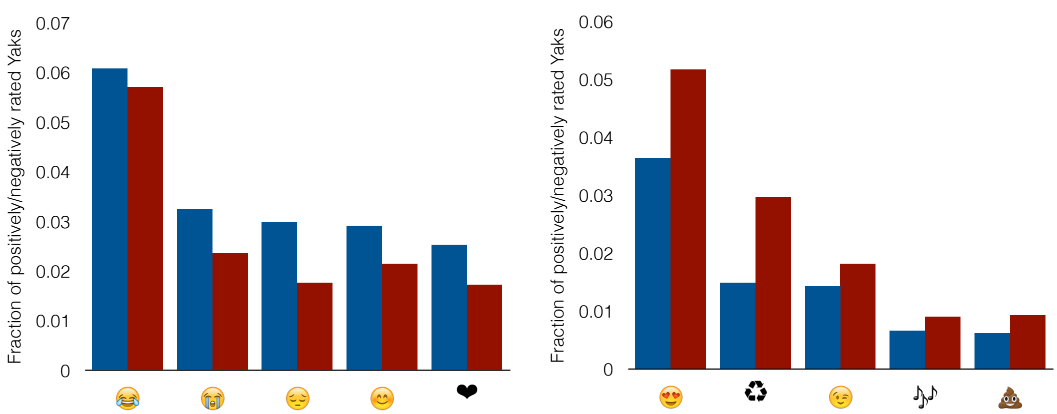

Figure 3: Emojis associated with upvoted (left) and downvoted (right) yaks. The bars show the frequency of the emoji in upvoted (blue) and downvoted (red) yaks (normalized by the total number of upvoted and downvoted yaks, respectively).

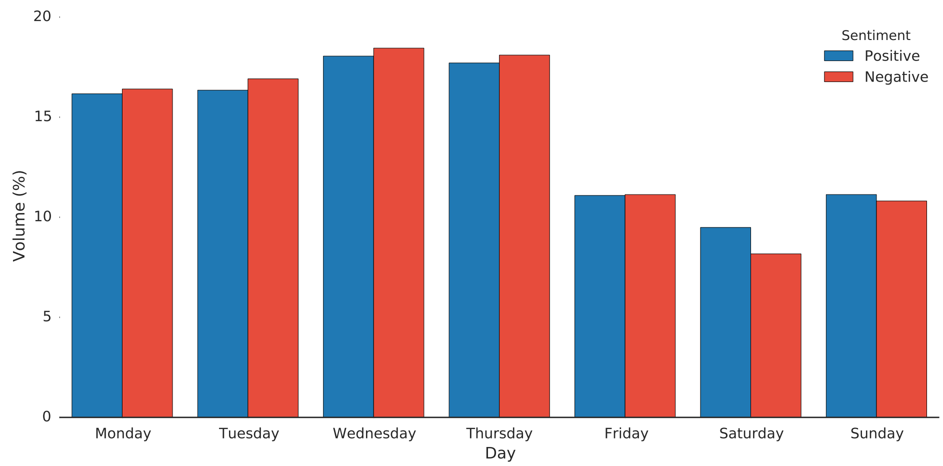

Figure 4: Volume of positive (blue) and negative (red) sentiment for different days of the week (normalized by the total number of positive and negative yaks, respectively). During weekdays yaks tend to be more negative, while during weekends they are more positive.

8 Sentiment Analysis

Classification. Next, we analyze the sentiment of yaks. We train a classifier with distant supervision, using emoji as labels for positive and negative sentiment. We labeled all yaks that contain: :-), :), :o), :], :3, :c), :), :D, =), 😃, 😄, 😊, 😸, 😻, 😎, as positive and all yaks that contain: :(, :-(, : (, :c, :[, 😁,😒,😖, 😞, 😠, 😡, 😢, 😩, 😪, 😫 as negative, resulting in a dataset of 47,537 labeled documents (23,437 positive and 24,100 negative). We experimented with different classifiers and weighting schemes and found that Naive Bayes with tf-idf weighting performs best. On 10-fold cross-validation, it achieves accuracy of 74.62% and AUC of 84.25%. The other classifiers: Logistic Regression, SVM, and Random Forest achieved similar results falling slightly behind. Finally, we trained the classifier on the full labeled (all yaks with positive or negative emojis) dataset and used the model to predict the remaining yaks.

Sentiment over time. Next, we analyzed how the sentiment changes over time. We plot the sentiment for different hours of the day and we find that it is mostly driven by the overall volume of yaks without any significant difference between the volume of positive and negative yaks during specific times of the day. However, when we change time scale from an hour to a day, we find something interesting (Figure 4). During weekdays the volume of negative yaks tends to be higher, on Fridays it is almost the same, and during weekends yaks tend to be more positive, likely due to the social patterns around universities.

9 Topic modeling

One of the key hypotheses regarding anonymous social networks 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 assigned each document (either a tweet or a yak) to the topics we derived to see if there was a strong bias between the two over topics. Using a Latent Dirichlet Allocation (LDA) model with 200 topics, we did not see a significant amount of topics that belonged heavily to one platform and not the other.

The sorted ratios of documents belong to topics 1-200 in either platform can be viewed in Figure 5. On average, topics tend to be evenly split across both platforms.

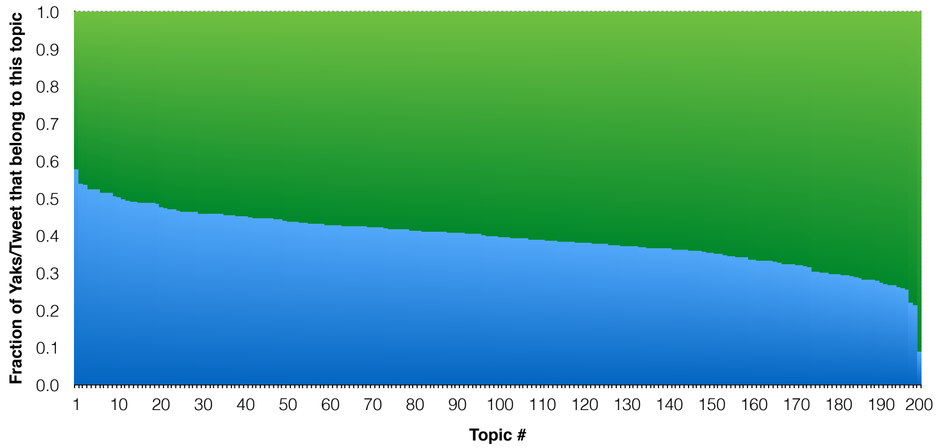

Figure 5: LDA topic modeling with 200 topics. The colors show the proportion of tweets (blue) and yaks (green) that belong to each topic. Most topics do not show bias towards either platform.

10 Conclusion

In our initial survey of behavior in anonymous social networks, we employ unprecedented empirical methods in attempts to verify assumptions surrounding communication. Most notably, we examine behavior on the social app Yik Yak, where users have no persistent identification and communication within a limited radius, with usage of Twitter (global and persistant identity) as a baseline.

Looking at temporal patterns, usage of vulgar vocabulary, frequency of emojis, discussion of topics, sentiment of statements and group mechanism such as community censoring, we have created a computational first view of key insights into anonymous social behavior.

Acknowledgements

We would like to thank Ivan Sysoev for sharing his implementation of the Mutual Information feature selection criterion, which provided the most interesting results, selecting both discriminative and frequent features, from all the measures we tried.

Appendix

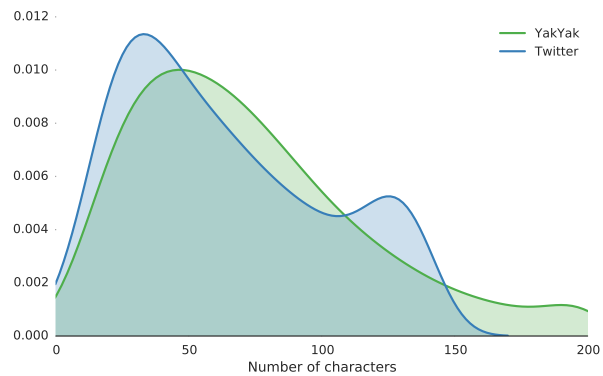

Figure 6: Distribution of posts length on Yik Yak (green) and Twitter (blue). Twitter has a limit of 140 characters, whereas yaks can be up to 200 characters long. This is reflected in the empirical distribution of the post length: on average yaks tend to be longer (72.49 characters) than tweets (63.91).

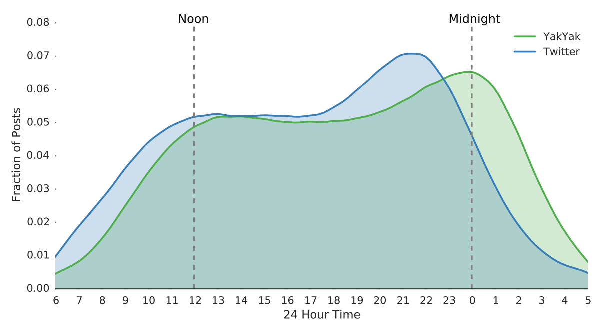

Figure 7: Posting activity on Yik Yak (green) and Twitter (blue) during different times of the day. Twitter users tend to post earlier in the day. The activity peak on Twitter is around 9pm (21h), while Yik Yak is most active around midnight (0h).

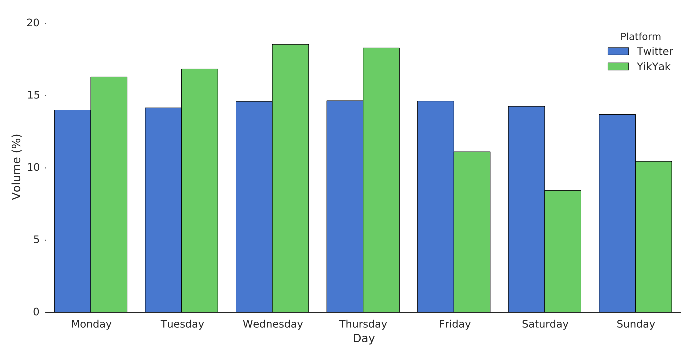

Figure 8: Posting activity during different days of the week. The volume of tweets is steady across the week, while there are more yaks during the weekdays (Wednesday and Thursday) and decline during the weekends.

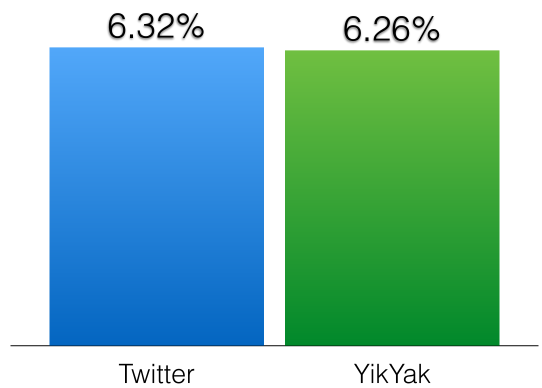

Figure 9: Fraction of tweets (blue) and yaks (green) that contain swear words.

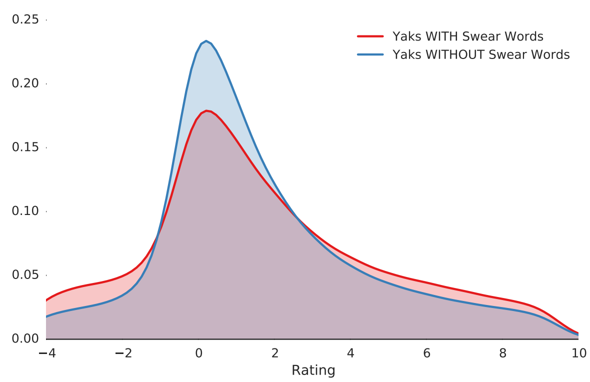

Figure 10: Distribution of ratings (upvotes - downvotes) on yaks that contain (red) and do not contain (blue) swear words.

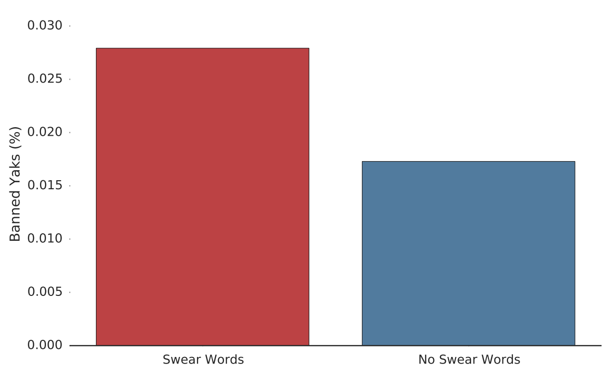

Figure 11: Fraction of banned yaks (with 4 downvotes) with and without swear words. yaks that contain swear words are 60% more likely to be banned.

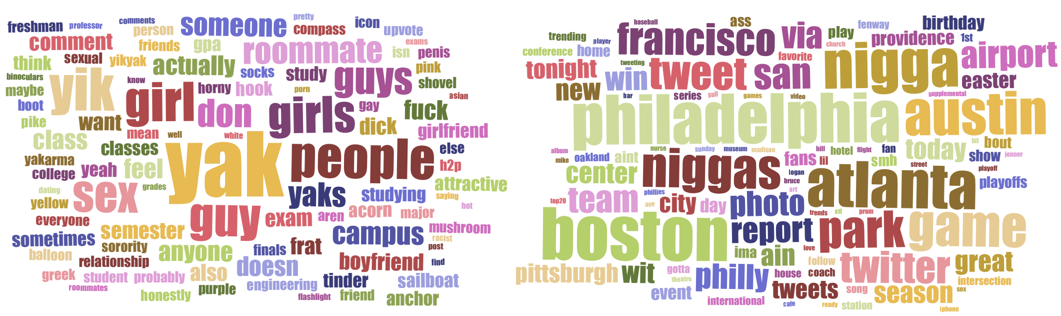

Figure 12: Wordclouds. Words characteristic for Yik Yak (left) and Twitter (right). We used Mutual Information as criterion to select the most discriminative words.