Tracking the Yak: An Empirical Study of Yik Yak
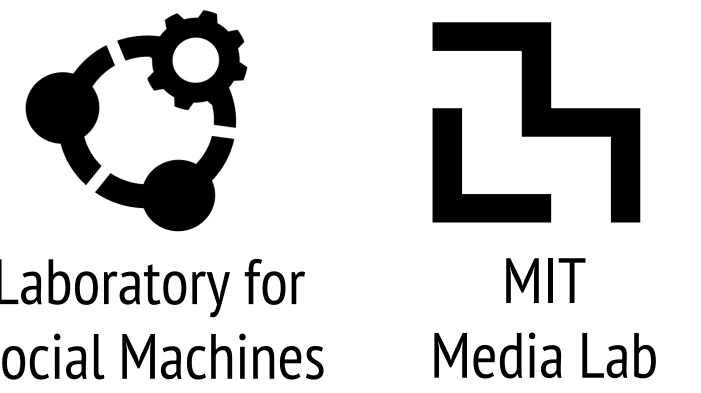
Martin Saveski, Sophie Chou, Deb Roy – Laboratory for Social Machines, MIT Media Lab

Laboratory for Social Machines

MIT Media Lab

ABSTRACT

To investigate the effects of anonymity on user behavior, we study the new and controversial social app, Yik Yak. As a comparison, we look at Twitter, which has similar limitations on lengths of posts, but is public and global rather than anonymous and local. 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.

THE YIK YAK APP

Yik Yak is a Twitter-like anonymous social smartphone app. Within a 10 mile radius, users can leave posts (yaks) of up to 200 characters and interact with other community member's posts. Users can upvote and downvote posts; if a post receives 5 downvotes, it is deleted, creating a self-selecting mechanism for censoring content.

DATA COLLECTION

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.

20 College Campuses in USA

19 Days

1.9M Yaks   1M Tweets

CONTENT ANALYSIS

Yaks vs Tweets. Yaks are dominated by words specific to college life, while tweets contain city names and sport related words.
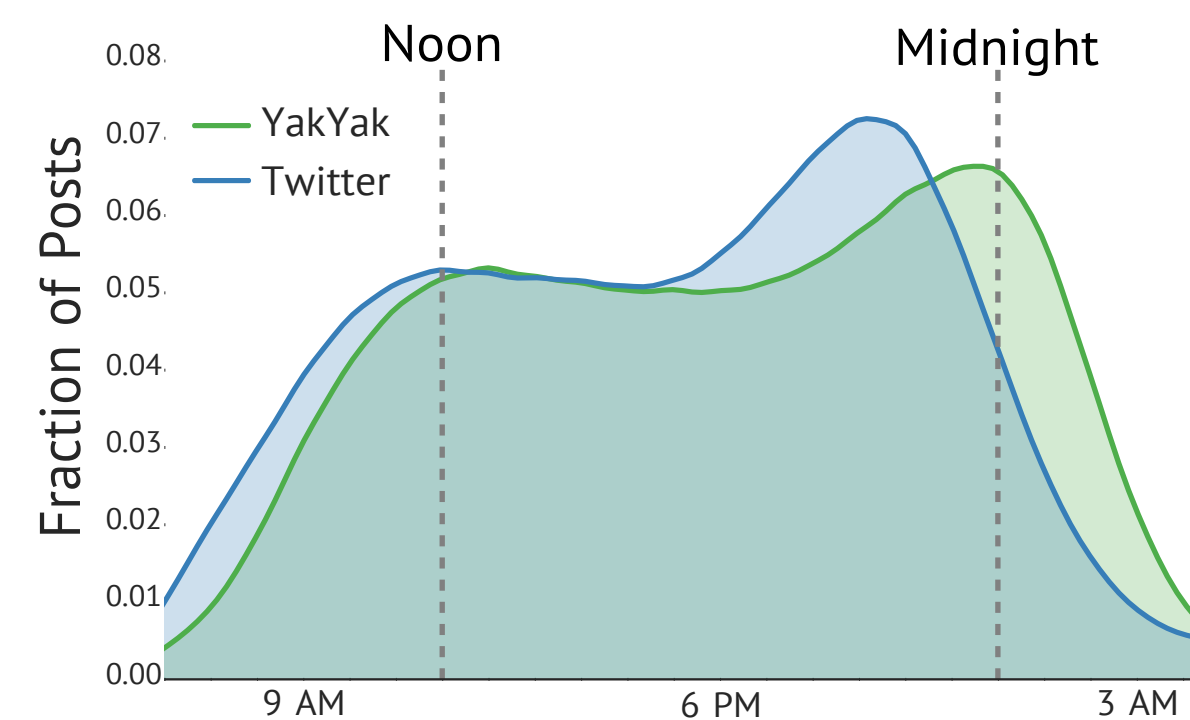

Yik Yak   Twitter
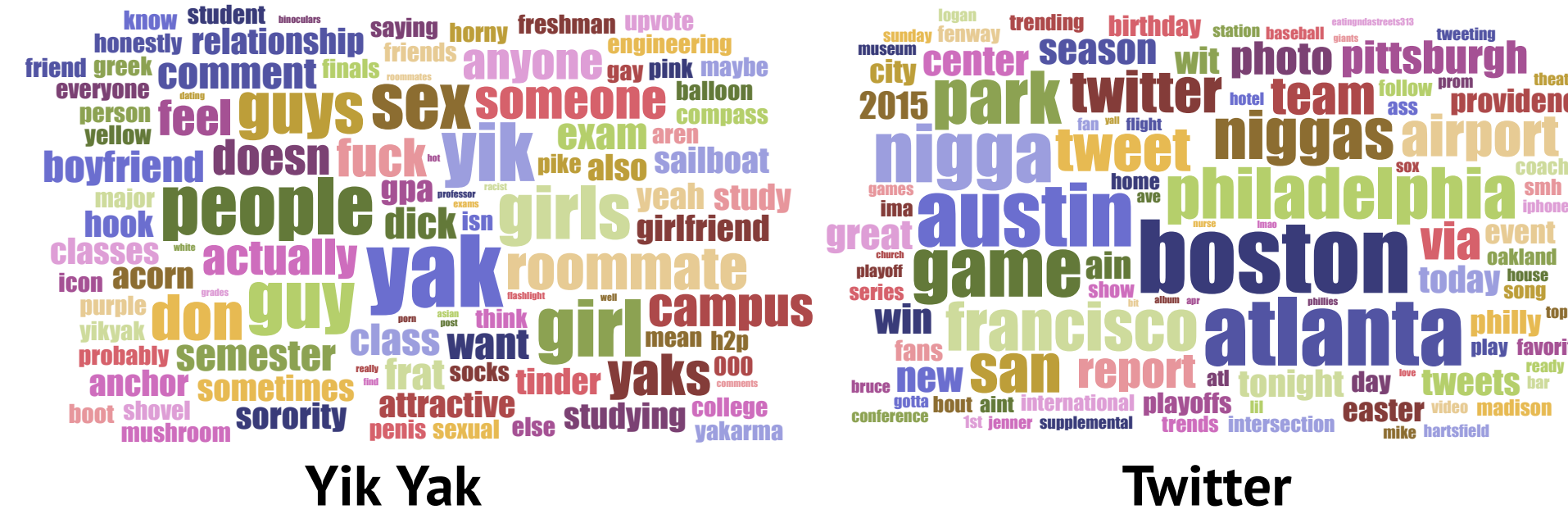
Upvoted vs Downvoted Yaks. Upvoted yaks are characterized by mostly general and slightly positive words, while downvoted yaks contains mostly inappropriate and racist words.
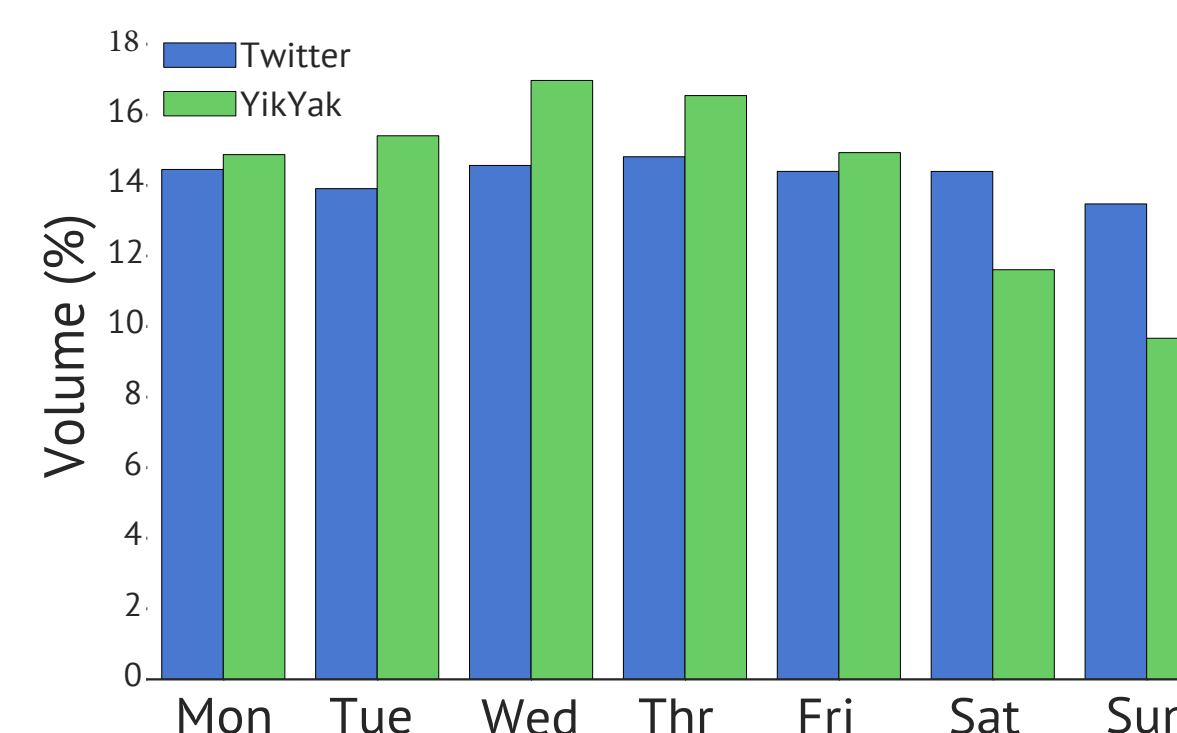

Upvoted   Downvoted
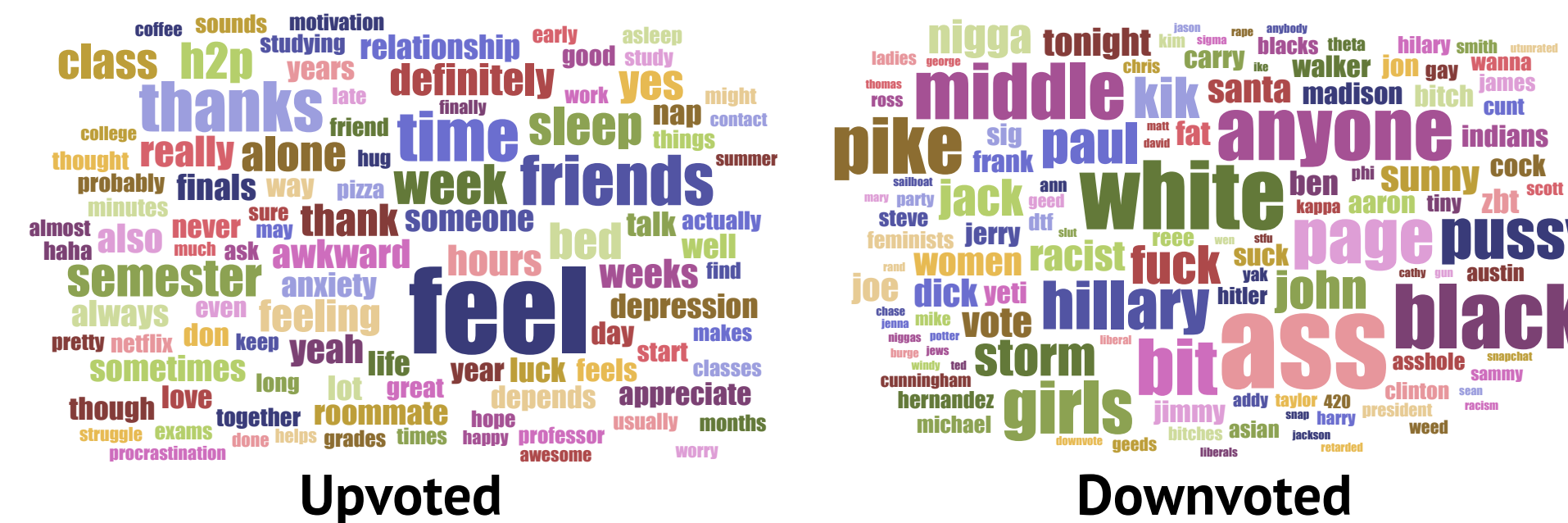
ACTIVITY PATTERNS

Daily Posting Patterns. Twitter users tend to post earlier in the day. The peak of activity on Yik Yak is at midnight, three hours later than on Twitter.
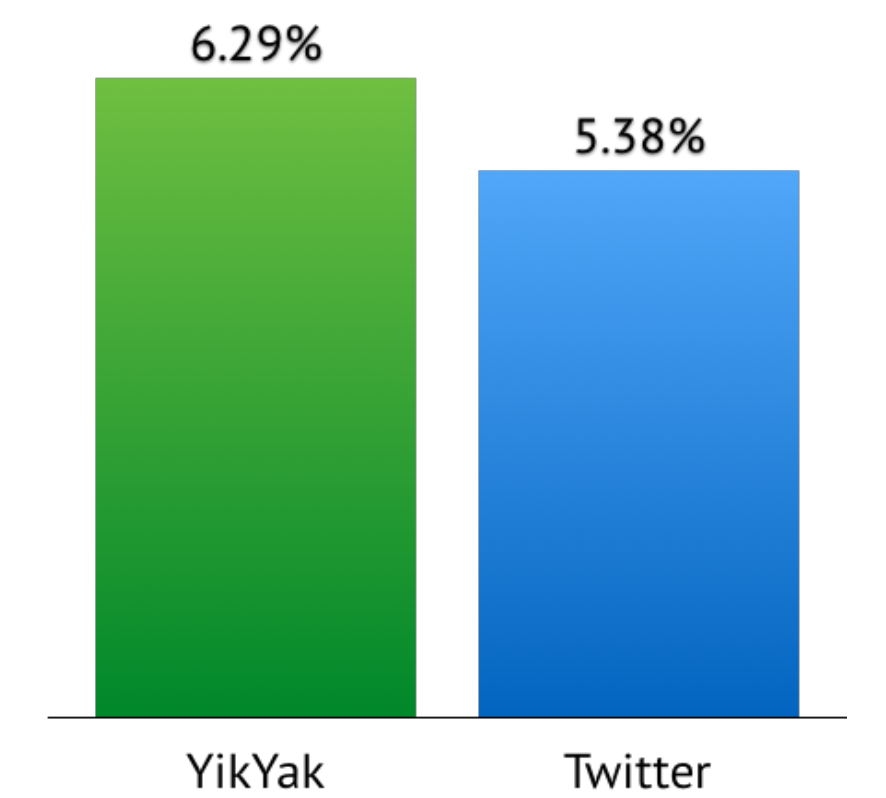

Weekly Posting Patterns. The posting activity on Yik Yak is highest on weekdays and declines during weekends. The volume of tweets is steady across the week.


USAGE OF VULGARITIES

RQ1: Is vulgar language more acceptable on Yik Yak than on Twitter?

Posts on Yik Yak are only slightly more likely to contain vulgar words than posts on Twitter. We find that yaks that contain vulgar words are 38% more likely to be downvoted and have negative ratings. Banned yaks are 61% more likely to contain vulgar words.

TOPIC MODELLING

RQ2: Do unique topics emerge in conversations on Yik Yak?
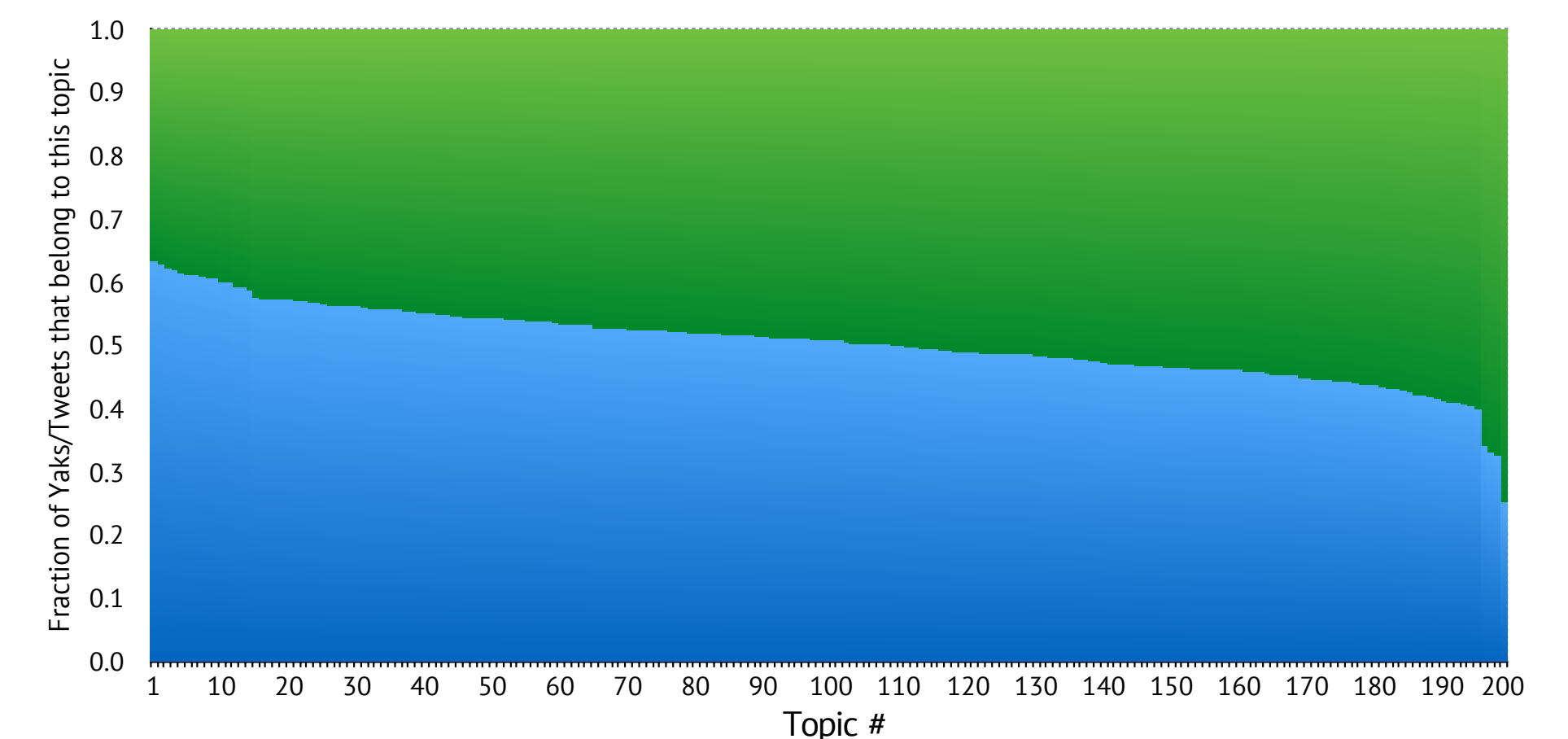
LDA topic modeling with 200 topics. The colors show the proportion of tweets (blue) and yaks (green) that belong to each topic.

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.

DISCUSSION

Limitations
• Missing content of banned yaks, they may contain unique topics,
• Geotagged tweets may not be representative of all tweets,
• Data from a short time period: 19 days.

Future Directions
• How topics vary across locations,
• Yik Yak as a social signal: are there any correlations between the activity on Yik Yak and the characteristics of the schools.