

Talking in the dark: An Analysis of the Anonymous Messaging Platform YikYak

Ramine Tinati, Dominic DiFranzo, Sergej Zerr, Markus Luczak-Roesch, Wendy Hall
University of Southampton
{r.tinati}@soton.ac.uk

ABSTRACT

Anonymous social networks are becoming increasingly popular in the age of Mobile and Web communications. They provide the same level of functionality as other well-known microblogging and social network services, yet promote the freedom for people to communicate and interact in an environment that seemingly preserves anonymity. However, despite the popularity of these services, relatively little is known about their usage and structure. In this paper we perform one of the first large-scale studies on the anonymous social messaging platform, *YikYak*. Based on a collection of over 500,000 Yaks and 1.6 million comments, we analyse *YikYak*'s structure, usage, and content. Our analysis reveals how individuals communicate, what topics they discuss, and how they make use of specific features to facilitate pseudo-anonymous conversations to take place. Based on our observations we compare the differences between anonymous and traditional social media platforms, and consider the wider social implications of their use.

Keywords

microblogging, anonymous, social media twitter, yikYak

Categories and Subject Descriptors

H.4.m [Information Systems Applications]: Miscellaneous

1. INTRODUCTION

For many years researchers have been interested in how we communicate, investigating the language we use and how our interactions differ between the private and public space (for example [22]). Since the advent of digital technologies such as the World Wide Web, Smartphones, and their *apps*, scholars are able to study this at unprecedented scale. From social networks, to microblogging services, to mobile messaging applications, we have witnessed a revolution in digital communication. Platforms such as Twitter and Weibo attract hundreds of millions of active users, and regularly boast figures of 500 million shared user-generated content, daily. Such platforms allow people to communicate with each other, to

broadcast text and multimedia content, to form 'friendships' or to follow each others' thoughts, emotions and mobility patterns.

Alongside the still growing number of platforms which provide communication and networking capabilities for users who can be persistently identified by their real name, a pseudonym, or at least their online account's URI, there has been an emerging class of social platforms and mobile applications, which claim to enable fully-anonymous interactions in a seemingly public space. These communication platforms allow individuals to interact and communicate without exposing any type of identifiable information or metadata. Examples exist among traditional online forums but also an increasing amount of mobile-only applications. Due to their system design, they offer the ability for people to say whatever they want, whenever they want, without being subjected to others knowing who they are. While this may be useful for some people to express themselves without the danger of damaging their reputation, which in itself can be problematic [14], it has also been shown that such applications can be valuable safe spaces for information exchange that may touch on sensitive socio-cultural topics such as sexuality [8, 9].

In response to the growing popularity and uptake of these platforms, there has been scholarly interest in understanding these new channels for anonymous communication. Research has been conducted to examine how the use of these anonymised communications platforms differs from services revealing the identity of a user (e.g. Twitter, Facebook, Reddit etc.) [39], considering not only their technical capabilities but also social trade-offs when operating within a private or a public space. A common motivation for studying these platforms relates to concerns over their potential misuse. Most recently, there has been a lot of mainstream attention on their role in antisocial behaviour; stimulating debate whether such tools should better be restricted or even forbidden because anonymised communications could be particularly harmful or problematic (e.g. anonymised hate speech).

Despite this increasing interest in studying these platforms, it is often technically difficult to collect data, or these services are too new to feature a relevant amount of user activity. Therefore, there is still little insight into human behaviour on those platforms and how they might change the social processes in the digital age. To fill this gap is the matter of the research presented in this paper.

We explore one of the fastest growing and most popular anonymous messaging applications, *YikYak*. Unlike other platforms, *YikYak* was designed as a *hyper-local microblogging platform*, which allows users to post anonymised messages only to an audience within their current geographic location, which in *YikYak*'s case is automatically linked to the nearest university and college

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

campuses¹. Our analysis, based on a corpus of over 500,000 regionally-diverse messages along with over 1.6 million comments, examines the emergent properties of content shared on the *YikYak* platform. To the best of our knowledge, this is the first large-scale, multi-region analysis of the *YikYak* platform.

1.1 Summary of Contribution

This work provides a large-scale exploratory platform study to the growing interest in understanding the use of large-scale social communication systems, with a particular focus on anonymous social messaging. It demonstrates subtle regional differences with respect to the discussions between users, specifically in the content and language style of the messages posted. In comparison to non-anonymous microblogging platforms such as Twitter, more discussion topics were found to relate to explicit content while the embedding of links to remote content is not common practice. Most significantly, our findings contribute to the scientific debate about anonymous communication in digital public spheres. It reveals techniques individuals deliberately use to de-anonymize information as well as behavioural patterns that may accidentally lead to de-anonymization, which motivates to revisit state-of-the-art assumption about disinhibition promoted by assumed anonymity.

2. BACKGROUND

There has been a rich history of studies which seeks to explore the tension between operating, interacting, and communicating in the private, vs public sphere [35, 36, 20]. Most prevalent in this area of research is the exploration of how humans interact and communicate when presented with an environment which does not expose their identity to other users or does not collect data about their identity. By comparing the types of conversations, the language used, and the context, it was found that anonymous environments lead to discussions which are much more controversial and less ‘political correct’ [21, 31].

Scholars have often drawn upon existing theories and frameworks to consider the reasons for why people use these anonymous platforms differently, with significant emphasis placed in understanding the role of the private space within the context of modern society [18]. Early work by Siegal et al. [35] and Suler et al. [36] developed the ‘Disinhibition’ effect based on empirical studies of early computer-mediated communications, and argues there are a number of behavioural characteristics at play between our online and offline interactions. In line with Siegal’s empirically-based studies, other theories have suggested that engaging in anonymous communication allows people to either expose their true underlying feeling, or establish another persona [25]. Anonymity encourages people to open up, especially those belonging to vulnerable parts of society [32, 15]. It can mask failure, and allow for more creative expression and experimentation [4]. Anonymity creates a safe space where people can reach out for help without fear of ridicule or abuse [30, 3]. Bodle [7] argues that anonymity helps to protect the fundamental rights of privacy and freedom of expression as declared in many constitutions and the Universal Declaration of Human Rights [1].

With respects to specific studies of anonymised messaging platforms and applications, research has examined various systems; from Web forums [34], to discussion boards [4], to question answering sites [17, 16]. Existing studies explore questions related to the content and the practises of users, with specific interests in understanding negative user behaviour, or behaviour which may

not be considered acceptable according to established social norms [34].

Pertaining to mobile-only applications, Wang et al. studied the Whisper social network platform [39]; their work represents one of the first large-scale studies of the structure and use of an anonymous messaging and networking platform. They explore the tracking and identification of users by exploiting the hidden location features which can be captured using traditional packet tracing methods. Similarly, Xue et al. [41] ran experiments to determine if it is possible to accurately identify users within the *YikYak* platform and found that using computational techniques, it is possible to identify the location of a Yak within a narrow radius. Namelka et al. [29] also performed similar exploitation analysis, finding that it was possible to identify specific users using the triangulation of different data sources.

With respects to the analysis of content and use of these platforms, Correa et al. [10] has explored the levels of anonymity sensitivity with respects to the content of messages within the Whisper platform, and can be accurately distinguished from platforms such as Twitter. Snapchat, one of the most recent mobile-only social media platforms has also been examined, with studies exploring social interactions and the types of content shared between users [33, 2].

Small-scale or geographically-restricted studies have been performed on *YikYak* such as Black et al. [5] analysis of 4,000 messages, which revealed how specific vocabulary and language can be observed in the messages. The only large-scale study of the *YikYak* platform has been conducted by McKenzie et al. [27], who examined a large corpus of Yaks from a single location, collected over a 5 month timeframe. The focus of their research was to explore the spatial aspect of Yakkers, and how it compares to other location-aware social services such as Twitter. Ma et al. [24] has shown that the types of communications and how individuals interact with each other are altered by the ability to mask one’s data and profile. They explore the concept of intimacy and self-disclosure, and their experiments reveal how as the level of anonymity increases, the level of disclosure increases, and that intimacy always regulates self-disclosure, independent of the environment. In an attempt to understand why people use these platforms, Kant et al. [19] conducted an interview-based study, revealing that people describe their participation is driven by being able to communicate and interact ‘honestly’ and ‘openly’.

In contrast to the existing work, the study presented in this paper seeks to build upon existing smaller-scale studies of the *YikYak* platform, and examine its structure, content, and usage. We also will contribute to the ongoing discussion regarding the type of user-generated messages and content published on anonymous platforms, and the wider societal implications and their role as modern social communication tools.

3. YIKYAK

YikYak is a mobile-only social media messaging application, available for various mobile platforms (Android and iOS). The purpose of *YikYak* is to enable anonymous messaging with others people in the same geographic location. In the initial version, *YikYak* assigned users (or ‘Yakkers’) are associated with a ‘campus’, which are situated around a university, college, or school campuses. Geofencing is used to separate different campuses, which at the time of writing, represents a geographic zone of 1.5 miles (radius).

Users are able to join their ‘local’ *YikYak* campus (based on their mobile phone GPS coordinates), however they are able to peek at messages from other campuses (but not contribute to). As shown in Figure 1, Users are able to construct messages (‘Yaks’), which are posted to a public campus timeline; all other users associated

¹YikYak currently only works in the United States and a few other countries.

with a *Campus* are able to read these messages. The Yak interface does not provide a rich set of features for writing a Yak, only plain text is allowed (e.g. Hashtags and URLs cannot be followed or clicked). By design, *YikYak* does not make it possible for users to scroll through a historic log of messages, they are only able to see the last few hours worth of messages, or less, depending on the volume of messages being produced. We assume that *YikYak* was designed this way in order to emphasize the real-time nature of discussion and conversation. All users remain anonymous, and users are not able to retrieve historic messages from the public timeline. To facilitate interaction between users and messages, *YikYak* provides a number of social voting and commenting features, including upvote/downvote, User metrics based on how active they are, and viewing ‘Top Yaks’, which are based on a set of features derived from the user profile and the ranking of Yaks. Yakkers are able to view and add comments associated with a given Yak, these are accessed via clicking on the Yak which then changes the mobile interface to display the comments and their associated metadata (votes, etc.). Comments also have their own associated metadata, however, nested comments are not allowed.

In April 2015, *YikYak* introduced the reply icons feature in order to allow for users to reply to a specific user within a thread of comments. As shown in Figure 2, these icons represent different objects, which are used as the profile image of a user within a thread. There is no specific order to these icons (they are randomly assigned), and they are not persistent beyond a single thread (e.g. a user can be assigned different icons in different threads). *YikYak* did not introduce a standardised way for users to direct their messages to a icon, so users adopted the @mention convention seen in other social platforms (e.g. ‘@socks’).

4. EXPERIMENTAL SETUP

The study in this paper is based on a corpus of 515, 704 Yaks collected during a two-week window in November 2015. We collected Yaks from 9, 705 US ‘campuses’, with their associated geographic location. For each Yak collected, *YikYak* assigns a school name and a given geographical Latitude and Longitude value. We validated that data against a published list of US schools and colleges, and also removed any Yak which not appear to have a valid location. Each Yak contained associated metadata including comments, rating (upvotes/downvotes), location, timestamp published, and a randomised identifier representing the user posting the message, as a consequence, it is not possible to perform user analytics (e.g. user feature extraction). Table 1 provides a general overview of the *YikYak* dataset.

Our aim is to compare the distribution of the conversation themes among the geographic regions presented through users and provide a bird’s eye view over this large corpora. For carrying out the experiments, we divided the Yaks of our data set into four different categories *Northeast*, *Midwest*, *West* and *South*, each corresponding to a geographic region of the United States that the Yak originated from. To identify the conversation themes we employ LDA (Latent Dirichlet Allocation) analysis [6] - the state of the art technique to determine latent topics in a document corpora used in different contexts. In our experiments we first compute such topics occurring in the whole corpora using LDA and then determine the representative topics for each of the given categories.

To this end, first by applying LDA we are able to represent latent topics as a list of terms with a probability for each term indicating the membership degree for the topic. Furthermore, for each *YikYak* in our corpus we can determine through topic probabilities $P(z_j|d_i)$ to which topics it belongs and to which degree. In order to identify representative topics we generalise a method, which

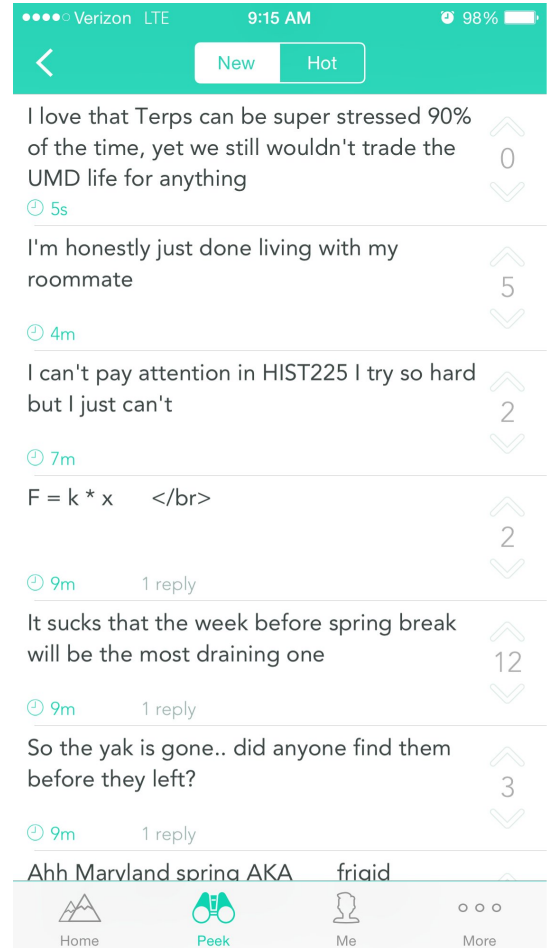


Figure 1: *YikYak* Mobile Application Interface

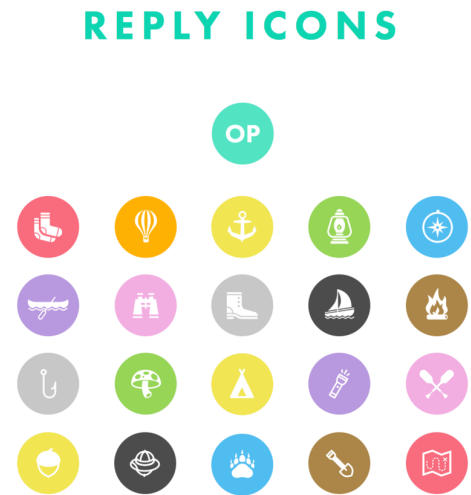


Figure 2: *YikYak* Profile Image Reply Icons

was previously employed in the context of the finding discriminative terms for text classification. Pointwise Mutual Information (PMI [26]) is an information theoretic measure to compute how much the joint distribution of topics and categories deviates from a hypothetical distribution in which topics and categories are independent of each other. We identify topics connecting two categories by combining the ranked topic lists from each of them using a harmonic mean.

Metric	Value
Yaks	515,704
Yak Comments	1,671,017
Corpus Start Date	2015-11-19
Corpus End Date	2015-11-30
Max/Min/Med/Avg Length of Yak	253/168/69
Max/Min/Med/Avg # of Comments	131/0/1/3
Max/Min/Med/Avg. # of URLs	3/0/0/0
Max/Min/Med/Avg. # of hashtags	12/0/0/0
Max/Min/Med/Avg Yak delta (secs)	720/0/100/96
Max/Min/Med/Avg comment delta (secs)	1500/0/168/172

Table 1: YikYak dataset overview

5. RESULTS

In this section we present a structural overview and associated descriptive statistics of the *YikYak* dataset, followed by the results of the topic modelling, separated by the different levels of location and geographic granularity (US States, and US regions).

5.1 Platform Analysis

Interested in the general usage of the *YikYak* platform, we analysed the general engagement patterns on the platform. After normalising all US campuses to the same timezone (EST), we found a diurnal pattern of activity with respects to the data/time at which Yaks are produced; As Figure 3 illustrates, activity tends to follow the hours of the students’ working hours, followed by a smaller peak in activity later in evening (20:00 - 24:00), which may correspond to evening social activities (the date/time were normalised to the timezone of different campuses). Such patterns of usage are similar to other social platforms and mobile applications (e.g. [28, 38]).

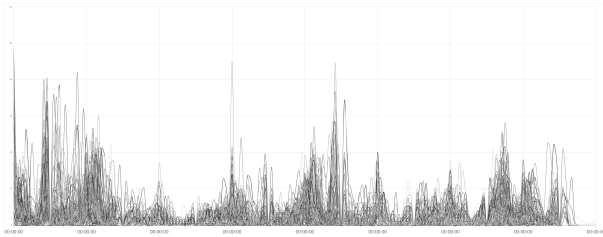


Figure 3: Daily Yak activity; timeseries represents the 100 most active campuses in each of the US Regions. Timeseries have been aligned to the EST timezone.

As shown in Table 1, despite no limitation to the length (characters) of a Yak, the average character length in 69 characters (median 60). We found this similar to other microblogging platforms such as Twitter [13] and Weibo, however these platforms are limited to 140 characters. We performed a similar approach as used by Li *et al.* [23], and implemented a simple lexical test to inspect the type of message published (e.g. a question, statement, or response). We found that 166,424 (32.3%) of Yaks were identified as questions,

and similarly, 238,685 (46.3%) of the comments were identified as questions.

With respects to the distribution of comments which a Yak received, we found similar proportions to other social communication platforms (e.g. Reddit, Whisper [40, 39]), where less than 3% of the Yaks received 95% of the total comments. On average, a Yak received 3 comments (median: 1), with the average length of a comment being 50 characters (median: 48), and the average length (chars) of comments for a Yak were 25% shorter in length (characters) than the initial message. With respect to the relationship between comments and Yak length, we found no significant relationship between the length of the Yak, and the number of comments received. With respects to the use of reply icons in comments, we found that 43% of comments contained a @mention to reply icon, and as Figure 4 illustrates 25.8% of these are referring to the ‘@OP’ icon, which refers to the Original Poster.

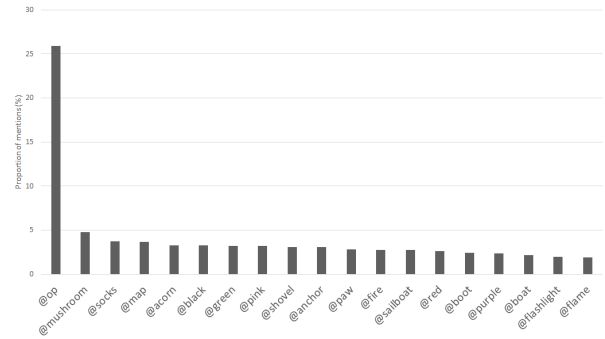


Figure 4: Number (%) of Reply Icon @mentions identified in Yak comments. A total of 2,941 @mentions were found. Chart shows the top 20 @mentions

We found that on average, a Yak (calculated at the campus-level, then averaged across all schools) would be published every 96 seconds, and for Yaks with comments, there were less than 3 minutes between the original Yak and the first comment being posted (and 60 seconds between comments). Our findings show that the first-to-comment response time is particularly fast, in comparison to Wang *et al.*[39] study of the Whisper platform where only 2.1% of the responses were within 60 seconds. Also, as an artifact of the platforms interface, we found that the decay in number of comments was fairly abrupt. The average delta between the first and last comment on Yaks with comments was 12 minutes, and we did not find any comment stream lasting longer than 30 minutes. One possible reason for this is due to the ephemeral nature of the discussions and messages posted in *YikYak*.

We also found very little engagement with the social voting features such as upvoting/downvoting, and the ability to ReYak (re-post) another message. Less than 5% of Yaks had voting metadata, and less than 0.1% of comments contained any voting information (a comment also contains metadata). This low level of engagement with social voting features is unlike other social platforms which provide social voting features (e.g. Reddit), which may be as a consequence of the ephemeral nature of the messages and the common practises of the *YikYak* users.

With respects to the aforementioned statistics regarding yak length, comments, and time between comments, We did not find any regional differences, and similarly, there was no statistical significance between the time of a Yak (e.g. during the day), and the length or number of comments received. An analysis of embedded resources in the Yaks found that less than 0.1% of Yaks contained a

URL, and only 0.08% of Yaks contained hashtags ('#'). Considering these findings, we assume that they are an artifact of the *YikYak* platform not explicitly supporting hashtags (i.e. a user is not able to click or follow a given hashtag in the *YikYak* interface), or automatic mechanisms to embed URLs (e.g. an explicit interface to share links to a specific Web page or resource).

5.2 Network Structure

YikYak does not explicitly support a typical social network structure; therefore it is not possible to follow or create friends within the platform (all Yakkers are anonymous). As a result, it is not possible to explicitly '@name' (mention) a specific Yacker within a message. In April 2015, *YikYak* introduced a feature which used a set of pre-defined profile pictures to represent different users within a comment thread. The platform randomly assigns these *reply icons*, thus each thread is unique, and there is no consistency between threads. Furthermore, although *YikYak* included these new reply icons, they did not include any explicit feature to allow users to directly mention the user with the assigned image. In response to this feature, users started to direct their message to the textual name of the reply image assigned to a user within a Yak comment thread (e.g. "hey @shovel, does it start at 2pm"). However, in addition to the @mentions to the specific reply icon (e.g. @paw), messages commonly contained the '@' symbol to comment about a person, place, or event ('@tod', '@superbowl', '@room123').

Interested in the @mention network, we constructed the social network between all Yaks and shared mentions. From the initial *YikYak* corpus, we extracted 3,287 edges, 1,667 nodes, consisting of US campuses and @mentions. As Figure 5 illustrates, we extracted a network structure between campuses, and identified common nodes which are not co-located geographically. For instance, *Campus A* (located in California) share the same '@mention' with another *Campus B* (located in Florida). As shown in both (a) and (b) in Figure 5, we visually identified this using different coloured edges between nodes. With respect to the connectivity of US Campuses, the average degree between campuses was 15.9, and the largest connected component (e.g. the largest network of campuses using a the same mentions) was 253. We found that by removing the '@mentions' to all profile image references, we were left with 89.5% of the total number of edges initially found. Examining the @mentions in these messages, identified many referring to a specific person (e.g. "@xxx is sitting in front of me"), or talking about a specific place or location (e.g. "I can't believe @starbucks is so busy").

We also constructed the @mention network by aggregating campuses into their representative US states. We did this for both the network with and without the reply icon @mentions. As Figure 6 illustrates, removing the reply icon @mentions, there are a number of commonly used @mention vocabulary which is consistent across different states. The average degree between states was 7.6 (and 11.2 with reply icons included), and the largest connected component (e.g. the largest network of connected states using a single mention) was 5 (with reply icons, the largest connected component was 49, connected by the '@OP' identifier). Inspecting the components, we found that certain states were connected by similar themes. For example, campuses in California, Florida, Ohio, and Georgia shared @game and '@kicker' (this formed the largest connected component in the graph); when examined, the content of the Yaks were discussing topics about their associated college football Team. We also found state specific vocabulary which was not not identified in Yaks outside a given state. For instance, '@cal-boyz' was found in messages from Californian US campuses, as was '@southernedu' in Connecticut.

5.3 Shared Topics

Region	Topic1	Topic2	Topic3
northeast	lol girl guy f*** kik	feel people good time love	class time good classes day
west	lol girl guy haha good	feel time love girl good	class good time day work
south	people black white make doesn't	class time home day work	Yak f*** lol yik sh*t
midwest	lol girl guy good feel	people make life doesn't white	class time good day classes

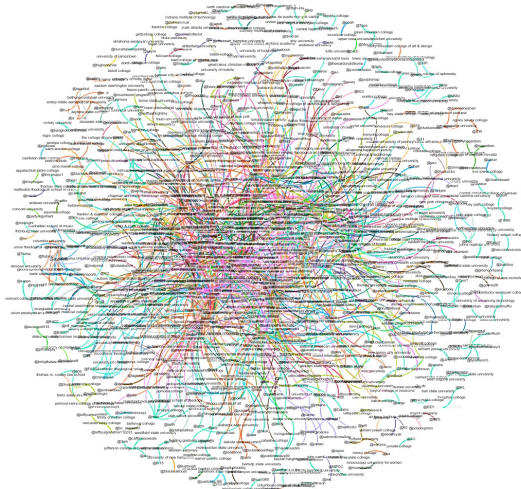
Table 2: Common Yak topics (and top 5 terms), grouped by US Region. Stop words not removed.

By using the LDA topic modelling technique as described previously in Section 4, we computed the represented set of shared topics at different data resolutions; school, state, and region (West, Midwest, Northeast, South). We found that due to the distribution of data collected for schools and states, aggregating topics at the level of the US region provide the most insightful results.

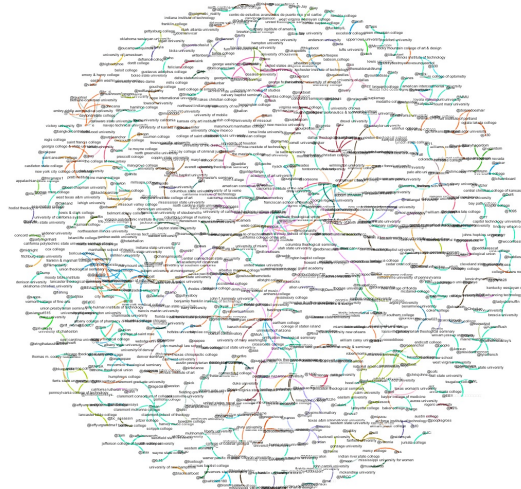
Figure 7 illustrates the distribution of the 200 most popular topics that were identified across all Yaks (including their comments) and are representative (with representativeness degree reflected in the font size) for each region. For this model, we developed a stop words list based on common profanities and slang. Additionally, we display the most common topics shared between different US regions, and the topics that are common for all the regions in the center of the illustration. For clarity the shared topics between non-adjacent regions (e.g. Northeast and West) have been hidden. Table 2 lists the top 3 topics for each of the four US regions (and the top 5 terms identified in each of these topics), based on a shared corpus of Yak messages. For this model, we did not remove any stop words; hence the extent to which profanities feature in the the topics. The classification of topics was performed by binning all Yaks from a given state into a chosen-US region. The mapping between the US states and US regions was based on the listing provided by the United States Census Bureau ².

The illustration of the shared topics in Figure 7 show that it is possible to identify regional differences based on the type of conversations which were taking place, along with a number of shared topics which could be considered as a national point of discussion; for example, national holidays politics, and work life. The topics identified in Figure 7 reflect the popularised views of lifestyle and behaviour across the different US regions. From instance, the West US region (in particular, California) is often considered more liberal, thus the identification of topics about 'smoking' and 'weed' would appear representative based on this popularised view. Similarly, the Midwest which is often known for US States which have long, cold winters and lots of snow, topics relating to the weather, are mostly representative. We also identified topics with Spanish words and phrases in the South US Region, which many reflect the demographics of users in the representative US States. Furthermore, the topics are representative at the point in time to which the data was collected (November, during the transition between fall and winter).

²United States Census Bureau, Geography Division. "Census Regions and Divisions of the United States".



a) YikYak (Full Network)



b) YikYak Network without Profile Label Edges

Figure 5: Bipartite network of US Campuses and shared @mentions found in Yaks and their comments. (a) illustrates the full network including the @mentions related to the profile pictures. (b) represents the network without the edges/nodes from the @mentions relating to the profile images (e.g. @shovel, @op), and contains 1,665 nodes, and 2,941 edges.

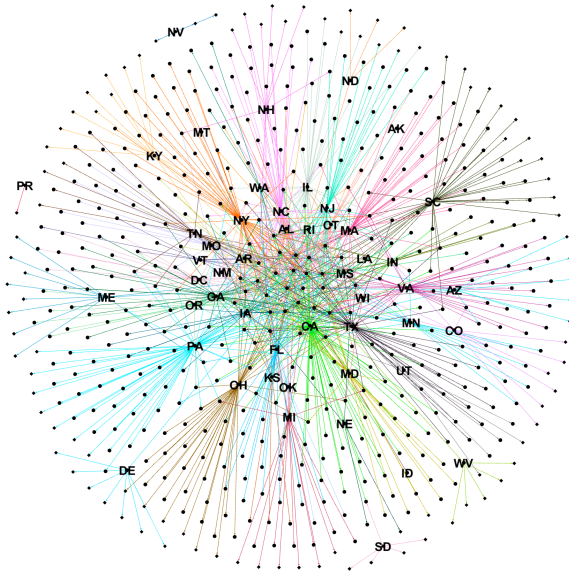


Figure 6: Bipartite network of US States (US Campuses aggregated to state level) and @mentions found in Yaks and their comments. All @mentions corresponding to profile picture (e.g. @op) have been removed. This network contains 587 nodes, and 2,336 edges.



Figure 7: Top representative topics (full text, including comments) distributed over the 4 US regions: West, Midwest, Northeast, South

Similar to similar studies of *YikYak* and other anonymous platforms [4, 5, 19], our analysis identified a strong emphasis of vulgar and abusive language and comments, which is illustrated clearly by the common topics shown in Table 2. Our observations reinforce the mainstream perception that anonymous social platforms facilitate communication which might not be suitable in a learning or educational environment, with schools now starting to ban its use by students. However, whilst many topics contained abusive language or demeaning content, our analysis identified topics (and via inspect, conversations) which were more positive in their content; these included conversations about learning, discussing school coursework, and offering advice (e.g. “go look at the slide 12 for CS121 if you’re stuck”).

Similar to other social networks where spam is heavily present from machine generated content (e.g. Twitter ‘bots’ [37, 11]), one could argue that *YikYak* suffers from this in the shape of abusive and vulgar messages. However, in contrast to more mature services such as Twitter, *YikYak* does not – or does not report to – filter or flag messages which may be spam (or adult-content). Such mechanisms may be suitable as the platform grows in the future, assuming they wish to attract users looking for high-quality, more relevant content that they can relate to [19].

6. DISCUSSION

Our analysis has revealed specific technical characteristics and social practises of the *YikYak* platform. Beyond the structural characteristics often observed in other platforms such as Twitter (e.g. diurnal patterns, skewed distributions of social comments to Yaks), *YikYak* reveals usage which is somewhat “darker” and more controversial in nature.

Most prominently, the topic analysis suggests *YikYak* is facilitating messages which would not be considered appropriate on traditional social networking platforms such as Twitter or Facebook. This is consistent with findings in other anonymous social networks and mobile apps like 4Chan, Secret, and Whisper. This is not to say that *YikYak* has no limits. All high schools in the US are geofenced and banned in *YikYak* to discourage high school students from using the app. Bomb threats and claims of violence are also strictly forbidden, and *YikYak* has worked directly with local and federal law enforcement to identify these individuals [12, 14].

Beyond the controversial characteristics of *YikYak*, there are also some interesting findings with regards to its users reusing customs and vocabularies from other online communities and platforms. The largest one being the ‘@’ mention. Our results showed widespread use of the ‘1@’ mention in relation to the reply icons that were developed for *YikYak*. The ‘@’ mention came to prominence from the early days of Twitter, and after its widespread use became a codified feature in that platform. But the ‘@’ mention is not an element expressed in the UI or rendering of Yaks on *YikYak*. *YikYak*’s users adapted this from their use and knowledge of Twitter. This unofficial and unsupported reuse of the ‘@’ mention has other interesting effects. Since there is no official or automatic naming convention for the Reply Icons that *YikYak* uses, local communities have developed their own. We identified many different vocabulary and conventions being used for the different reply icons (e.g. ‘@purplesocks’, ‘@socks’, ‘@purple’). Whilst we did not examine this in-depth, we found that certain campuses have adopted a given term or vocabulary, and have become the norm. For example in a Californian University campus, the flashlight icon was consistently used as ‘@flashy’, and was in many Yak threads. Whilst the result of this means that there has been some social-protocol reached for using a shared vocabulary, it could also mean that newcomers can

be easily identified, as they do not know the normal social practises of the community.

Another example of this reuse social practises from other platforms can be found in the ‘@OP’ identifier. This term, which stands for *Original Poster*, comes from the early days of web forums and grew in popularity in its use in communities like 4chan and Reddit. This identifier was used by *YikYak* users, and then formally adopted by the *YikYak* platform through the implementation of the Reply Icons.

Likewise just because *YikYak* users know of and reuse conventions of other platforms, does not mean they use all of these conventions wholesale. For example, hashtags are a popular way to categorise a message to a larger conversation on Twitter and Facebook. This is another convention that started informally by Twitter users (as was the ‘retweet’ RT @screenname function), and was later added into the Twitter platform. But our results show that hashtags are not used in Yaks. This may be because of the hyper-local ephemeral nature of the platform, or simply because the platform does not explicit support linking or following hyperlinks (on Twitter, hyperlinks were always able to be followed). Our results show that Yaks do not stay in a timeline long; and interaction takes place with a Yak in less than 3 minutes from its post, and these interaction (comments on the Yak) typically last only 12 minutes. You cannot search on *YikYak*, so categorising a Yak does not make sense if it’s just going to disappear as soon as more messages are created.

What we are observing is the *YikYak* community finding solutions from outside communities and platforms for issues that they have, which in this case is facilitating discussions in a comment thread. What makes this even more interesting, is that there is not a single *YikYak* community in the traditional sense. *YikYak* by its nature is hyper-local, users can only interact to their physically local community. These small communities not only find solutions to their problems, but they are finding the same solution. This can be also be seen in the existence of an implicit social network through the shared uses of mentions. We found a network which existed between campuses and US states, and the connected components which emerged showed relevance and context to the US states or Campuses which they connected. We believe this may be happening because they are a part of these larger, globally connected, online communities (like Facebook, Twitter and Reddit).

These observations may reveal aspects of the type of users engaging with the *YikYak* platform; rather than attracting a new set of novice users, there may be a migration of users from existing services. This would contradict the mainstream view that these new breed of services, such as *YikYak*, are enabling a new set of users to become digitally and socially engaged [39]. Moreover, based on this argument, it might be possible to infer the demographics of the users, and help explain why the services are being used (or attracting) specific forms of interaction.

It has been argued that anonymity promotes the disinhibition effect in online environments [34], that being, the behavioural trait of social norms which are typically followed in an offline space are ignored or forgotten online [35, 36]. However, unlike services and platforms of a previous generation (e.g. forums and discussion boards), location-based mobile applications have the capacity to reduce the gap between the online and offline. The users of a service such as *YikYak* are locally close to other participants, and, based on the interactions we have observed, are interacting in identifiable, or pseudo-identifiable ways. If the disinhibition effect remains true with the use of these new services, then we must consider the social implications of this.

7. CONCLUSION

In this paper, we conducted one of first large-scale studies of the anonymous social messaging platform, *YikYak*. Our study explores the structure, usage, and content of the platform, and considers its role as a social messaging application. Our findings contribute to the scientific debate about anonymous communication in digital public spheres[36]. It reveals techniques individuals deliberately use to de-anonymize information as well as behavioural patterns that may accidentally lead to de-anonymization, which motivates to revisit state-of-the-art assumption about disinhibition promoted by assumed anonymity.

Our analysis show that people make use of *YikYak*'s ephemeral nature, with short bursts of discussion, and relatively little engagement on messages which are historic. We found a lack of engagement with the social voting features, which is likely an artifact of the how messages are part of a real-time stream of content, and do not remain present for a significant amount of time (and there are no historic records for access). We found users tend to communicate about topics which would be considered vulgar and abusive on traditional social platforms. Whilst this is not surprising (as studies in of other anonymous environments have shown similar conversations), what's interesting is how individuals' are using alternative means to identify and communicate with other users. In addition to the use of profile image @mentions (which were introduced by *YikYak* in response to users not being able to address their comments to others users within a message thread), individuals are being identified by common vocabulary, which extend beyond the system-enforced geographic localities. Furthermore, despite *YikYak* being a location-based service, we identified a significant amount of shared topics and conversations across the different geographic areas explored (US campuses and states). Whilst we did identify regional-specific conversations (e.g. the Midwest campuses talking about snow), there was a significant amount of cross-over in terms of the general discussions, with shared topics related to studying, partying and general student life.

Considering *YikYak* and its societal role in social messaging, similar to other studies of anonymous services (e.g. Whisper, Secret), our findings suggest that these class of services provide an alternative back-channels of communication, which may feature discussions which may be considered as not appropriate in a given environment (e.g a classroom or school). Moreover, Given that these services give rise to discussions which could be considered abusive, future design of these platforms may consider how filtering, or mechanisms to stop the mentioning or pseudo-mentioning of individuals could be performed. This is particularly important where applications are location-based, anonymity maybe jeopardised.

In our future work, we wish to explore two avenues; investigating the transfer of platform-specific practises and usages, and how these extend and transfer between different social platforms, seeking to understanding how these practises then develop and shape the usage and design of the system. For example, how did the introduction of the '@OP' identifier, which itself was originally a socially-constructed mechanism from existing Web communities, alter the usage of the *YikYak* platform? Understanding this process of system-to-system knowledge transfer will offer insight into how platforms such as *YikYak* are developing, and their potential future role, integration in as social communication tools.

Acknowledgement

Acknowledgements have been removed for review purposes.

References

- [1] Assembly, U. G. Universal declaration of human rights. *UN General Assembly* (1948).
- [2] Bayer, J. B., Ellison, N. B., Schoenebeck, S. Y., and Falk, E. B. Sharing the small moments: ephemeral social interaction on snapchat. *Information, Communication & Society* 19, 7 (2016), 956–977.
- [3] Baym, N. K. *Personal connections in the digital age*. John Wiley & Sons, 2015.
- [4] Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., and Vargas, G. G. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In *ICWSM* (2011), 50–57.
- [5] Black, E. W., Mezzina, K., and Thompson, L. A. Anonymous social media - understanding the content and context of yik yak. *Comput. Hum. Behav.* 57, C (Apr. 2016), 17–22.
- [6] Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *The Journal of machine Learning research* 3 (2003), 993–1022.
- [7] Bodle, R. The ethics of online anonymity or Zuckerberg vs. Moot. *ACM SIGCAS Computers and Society* 43, 1 (2013), 22–35.
- [8] Bosch, T. Wots ur ASLR? Adolescent girls' use of cellphones in Cape Town. *Commonwealth Youth and Development* 6, 2 (2008).
- [9] Chigona, W., Chigona, A., Ngqokelela, B., and Mpofu, S. Mxit: Uses, perceptions and self-justifications. *Journal of Information, Information Technology, and Organizations* 4, 1 (2009), 16.
- [10] Correa, D., Silva, L. A., Mondal, M., Benevenuto, F., and Gummadi, K. P. The many shades of anonymity: Characterizing anonymous social media content. *Proc. of ICWSM 15* (2015), 71–80.
- [11] Gilani, Z., Wang, L., Crowcroft, J., Almeida, M., and Farahbakhsh, R. Stweeler: A framework for twitter bot analysis. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, International World Wide Web Conferences Steering Committee (Republic and Canton of Geneva, Switzerland, 2016), 37–38.
- [12] Glum, J. In growing trend, yik yak users post anonymous school bomb threats – and get caught. *International Business Times* (nov 2014).
- [13] Go, A., Bhayani, R., and Huang, L. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford 1* (2009), 12.
- [14] Guardian. Stanford, the swimmer and yik yak: can talk of campus rape go beyond secrets?, February 2015. [Online; posted 8-February-2015].
- [15] Henderson, S., and Gilding, M. "I've never clicked this much with anyone in my life": Trust and hyperpersonal communication in online friendships. *New Media & Society* 6, 4 (2004), 487–506.

- [16] Hosseinmardi, H., Ghasemianlangroodi, A., Han, R., Lv, Q., and Mishra, S. Analyzing negative user behavior in a semi-anonymous social network. *CoRR abs/1404.3839* (2014).
- [17] Hosseinmardi, H., Ghasemianlangroodi, A., Han, R., Lv, Q., and Mishra, S. 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 (2014), 244–252.
- [18] Kang, R., Brown, S., and Kiesler, S. 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 (2013), 2657–2666.
- [19] Kang, R., Dabbish, L., and Sutton, K. Strangers on your phone: Why people use anonymous communication applications. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16*, ACM (New York, NY, USA, 2016), 359–370.
- [20] Kerr, I. R., Steeves, V. M., and Lucock, C. *Lessons from the identity trail: anonymity, privacy and identity in a networked society*. Oxford University Press, USA, 2009.
- [21] Kiesler, S., Siegel, J., and McGuire, T. W. Social psychological aspects of computer-mediated communication. *American psychologist* 39, 10 (1984), 1123.
- [22] Klapper, J. T. The effects of mass communication.
- [23] Li, B., Si, X., Lyu, M. R., King, I., and Chang, E. Y. Question identification on twitter. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, ACM (2011), 2477–2480.
- [24] Ma, X., Hancock, J., and Naaman, M. Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM (2016), 3857–3869.
- [25] Maczewski, M. Exploring identities through the internet: Youth experiences online. In *Child and Youth Care Forum*, vol. 31, Springer (2002), 111–129.
- [26] Manning, C. D., Raghavan, P., and Schütze, H. *Introduction to Information Retrieval*, 1 ed. Cambridge University Press, July 2008.
- [27] McKenzie, G., Adams, B., and Janowicz, K. Of oxen and birds: Is yik yak a useful new data source in the geosocial zoo or just another twitter? In *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN'15*, ACM (New York, NY, USA, 2015), 4:1–4:4.
- [28] Naaman, M., Zhang, A. X., Brody, S., and Lotan, G. On the study of diurnal urban routines on twitter. In *ICWSM* (2012).
- [29] Nemelka, C. L., Ballard, C. L., Liu, K., Xue, M., and Ross, K. W. You can yak but you can't hide. In *Proceedings of the 2015 ACM on Conference on Online Social Networks, COSN '15*, ACM (New York, NY, USA, 2015), 99–99.
- [30] Nissenbaum, H. The meaning of anonymity in an information age. *The Information Society* 15, 2 (1999), 141–144.
- [31] Nussbaum, M., and Levmore, S. The offensive internet: Speech, privacy and reputation.
- [32] Qian, H., and Scott, C. R. Anonymity and self-disclosure on weblogs. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1428–1451.
- [33] Roesner, F., Gill, B. T., and Kohno, T. Sex, lies, or kittens? investigating the use of snapchat's self-destructing messages. In *International Conference on Financial Cryptography and Data Security*, Springer (2014), 64–76.
- [34] Schoenebeck, S. Y. The secret life of online moms: Anonymity and disinhibition on youbemom. com. CiteSeer (2013).
- [35] Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. W. Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes* 37, 2 (1986), 157 – 187.
- [36] Suler, J. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.
- [37] Thomas, K., Grier, C., Song, D., and Paxson, V. Suspended accounts in retrospect: an analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, ACM (2011), 243–258.
- [38] Tongaonkar, A., Dai, S., Nucci, A., and Song, D. Understanding mobile app usage patterns using in-app advertisements. In *International Conference on Passive and Active Network Measurement*, Springer (2013), 63–72.
- [39] Wang, G., Wang, B., Wang, T., Nika, A., Zheng, H., and Zhao, B. Y. Whispers in the dark: analysis of an anonymous social network. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, ACM (2014), 137–150.
- [40] Weninger, T., Zhu, X. A., and Han, J. An exploration of discussion threads in social news sites: A case study of the reddit community. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, IEEE (2013), 579–583.
- [41] Xue, M., Ballard, C., Liu, K., Nemelka, C., Wu, Y., Ross, K., and Qian, H. You can yak but you can't hide: Localizing anonymous social network users. *ACM Internet Measurement Conference (IMC) 2016* (2016).