What's Kooking? Characterizing India's Emerging Social Network, Koo

Asmit Kumar Singh*§, Chirag Jain*§, Jivitesh Jain^{†§}, Rishi Raj Jain*§,

Shradha Sehgal^{†§}, Tanisha Pandey^{*§}, Ponnurangam Kumaraguru^{†#}

*Indraprastha Institute of Information Technology, Delhi, India

[†]International Institute of Information Technology, Hyderabad, India

 $\{asmit18025, chirag17041, rishi18304, tanisha17116\} @iiitd.ac.in \\$

 $\{jivitesh.jain,\ shradha.sehgal\} @ students.iiit.ac.in,\ pk.guru@iiit.ac.in \\$

Abstract-Social media has grown exponentially in a short period, coming to the forefront of communications and online interactions. Despite their rapid growth, social media platforms have been unable to scale to different languages globally and remain inaccessible to many. In this paper, we characterize Koo, a multilingual micro-blogging site that rose in popularity in 2021, as an Indian alternative to Twitter. We collected a dataset of 4.07 million users, 163.12 million follower-following relationships, and their content and activity across 12 languages. We study the user demographic along the lines of language, location, gender, and profession. The prominent presence of Indian languages in the discourse on Koo indicates the platform's success in promoting regional languages. We observe Koo's follower-following network to be much denser than Twitter's, comprising of closely-knit linguistic communities. An N-gram analysis of posts on Koo shows a #KooVsTwitter rhetoric, revealing the debate comparing the two platforms. Our characterization highlights the dynamics of the multilingual social network and its diverse Indian user base.

Index Terms—Social Computing, Data Mining, Online Social Media, Computational Social Science, Koo App, Twitter

I. INTRODUCTION

With 4.66 billion users worldwide,¹ the Internet has become a mainstream medium for social interaction and information dissemination. Social media is an integral part of the Internet, as it enables users and organizations worldwide to connect, socialize, and express themselves with ease to a large audience. With 560 million Internet users, India ranks second in the world. However, English is the first language for only 0.02% of the Indian population,² thereby creating a language barrier for the majority of the population [1]. The diverse set of languages

Major part of this work was done while Ponnurangam Kumaraguru was a faculty at IIIT-Delhi.

[§]Equal contribution. Arranged in the alphabetical order of the first name. ¹https://bit.ly/3d2VzRB Accessed 11 March 2021.

²https://bit.ly/3vtwWno Accessed 11 March 2021.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '21, November 8–11, 2021, Virtual Event, Netherlands © 2021 Association for Computing Machinery. ACM ISBN 978-1-4503-9128-3/21/11...\$15.00 http://dx.doi.org/10.1145/3487351.3488354 spoken in India raises the need for multilingual social media platforms in Indian languages.

Recognising this need, Koo is a recent attempt at building a multilingual Indian social network. Formerly known as *Ku Koo Ku*, it was launched in March 2020 as an Indian alternative to the popular social networking service, Twitter. Originally available in Kannada, Koo allows for and encourages discourse in Indian languages and currently supports 9 other Indian languages apart from English.³

Koo gained popularity during August 2020, when it won the Government of India's Aatmanirbhar Bharat App Innovation Challenge Award.⁴ During the Indian Farmers' protest in January 2021, Twitter entered a week-long standoff with the Indian Government, over their refusal to block accounts that the Indian Government claimed were spreading misinformation.⁵ Consequently, several Indian Government ministers, officials and agencies⁶ created accounts on Koo. There has been a general promotion of Koo amongst Government organizations since this event, sparking a rise in the platform's popularity. It again saw an increase in the influx of users when Koo became the first platform to agree to abide by the Government of India's Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021⁷, while Twitter and Facebook resisted them. Koo has even announced its entry into Nigeria, with the Nigerian Government creating an account on the platform, following the ban of Twitter in the country.⁸

Koo uniquely positions itself as an alternative to mainstream online social networks like Twitter. A cardinal reason for its popularity in the Indian context is its multilingual support that allows for more inclusivity. The language barrier is broken, and more people can now join social media and express and share their opinions in the language that they choose. Additionally, Koo has received support from the Indian and Nigerian government agencies, contributing to its acclaim. With the ever-increasing political discourse on social media, Koo becomes an important platform from a political standpoint.

⁵https://bit.lv/3zwWaUS Accessed 21 March 2021.

⁸https://bit.ly/2RXEaCi Accessed 14 June 2021.

³At the time of data collection, support for them was under development. ⁴https://bit.ly/2RWGI3w Accessed 12 March 2021.

⁶https://bit.ly/3xuECY1 Accessed 21 March 2021.

⁷https://bit.ly/3AUkwZh Accessed 21 March 2021.

The rich diversity of Indian languages on the platform coupled with its sudden rise and steady expansion in popularity and political backing motivate the need to understand what goes on inside such a unique platform. To understand this, we provide, to the best of our knowledge, the first characterization of the Koo social network. We do this by investigating the following research questions:

- 1) **RQ1:** What are the characteristics and demographic of Koo users? When did they join the platform?
- 2) **RQ2:** What kind of content is posted on the platform? Which languages are most popular?
- 3) **RQ3:** What are the network properties of Koo and how do they differ from Twitter's? What communities are formed on the platform?

We were able to collect data for 4.07 million users out of the total 4.7 million⁹ users and 163 million follower-following edges in our data (Section III). In Section IV, we analyse the user demographic - we found that even though the users on the platform are predominantly male, female reported users are more active, and have more average followers and average likes. We also found that the influx of users surged on the platform in August 2020 when it won the Atmanirbhar Bharat App Innovation Challenge, and in the starting months of 2021 when the government promoted the App. On analysing the content in Section V, we saw that the most popular language on Koo is Hindi, followed by English, Kannada and Telugu. We also observed a Koo vs Twitter rhetoric on the platform and support for the Indian political party BJP. We saw that major political figures were one of the most mentioned users on Koo. On comparing the network properties of Koo and Twitter (Section VI), we found Koo to have a dense, well connected network with a higher clustering coefficient. Furthermore, we observed distinct communities of users on Koo, which are based on languages, with English speaking users more centrally placed and having connections to Hindi and Kannada users.

Through our work, we make the following contributions:

- 1) Perform an extensive characterization of the new Indian social network Koo, in terms of its user demographic and content.
- 2) Present the first dataset of users, their connections, and content on Koo.
- 3) Study the network and communities formed on this multilingual platform.

II. RELATED WORK

In this section, we review previous work on social media analysis, in particular the multilingual nature of platforms.

There has been a vast amount of research on popular social media, Twitter. This includes its characterization in its initial years [2] and the documentation of its steady user growth [3]. With the growth of the platforms, the amount of research increased as well, with new topics of interest emerging, such as fake news and misinformation [4] [5], automated bots [6], hate speech [7] [8], etc. Some works have explored Indian

⁹As of 16th March 2021, https://bit.ly/3d2VcGH

languages on Twitter, with many using it as a corpus of indic-NLP research [9] [10] or studying the topics of discourse in the Indian subcontinent [11] [12]. Our work differs substantially from these as we focus on a platform dedicated to promoting Indian languages, with a predominantly Indian user base.

There have been studies exploring the emergence of new social platforms with unique properties such as Whisper for its anonymity [13], TikTok for its short videos [14], Twitch for mixed media [15], and so on. Some Twitter alternative alt-right platforms like Gab [16] [17] and Parler [18] have also been analysed in the past. Other social media platforms in local languages have also been studied, like VKontakte in Russia [19] and Weibo in China [20].

Multiple studies look at the multilingual aspect of other social platforms as well like blogs [21] and reviews [22]; but most of these cases find the English language to be dominant over the other languages. Our study focuses on a platform that has a native language to be dominant. Agarwal et al. [23] do study the qualities of an Indian multilingual social network, Sharechat. However, it focuses mainly on image-based content posted by the user. Koo is majorly a text-based platform, thereby making an inquiry of language aspects even more pertinent. Apart from the content posted, our work also lays emphasis on the user characteristics and network of Koo and draw a comparison of the platform's properties to that of Twitter. The choice of a novel platform and the context behind its rapid rise in popularity differentiates our work from the remaining.

III. METHODOLOGY

We describe next our data collection methodology. We start by presenting an overview of the platform.

A. Overview of Koo platform

Similar to Twitter, Koo allows logged-in users to share microblog posts known as "*koos*". While signing up, users can choose their display name, handle, language and other personal details such as gender, marital status, and birth date. Once logged-in, users can post koos, which can at most be 400 characters long; whereas, Twitter allows tweets to be 280 characters at maximum. Users can also comment on others' koos or re-share (also known as "*rekoo*") them with or without adding a comment. Koo provides transliteration support while typing in native Indian scripts. The platform's user interface, the users' feed, and the list of recommended accounts to follow changes according to their chosen language, enabling the formation of linguistic communities on the platform.

B. Data Collection

We collected data pertaining to the users' profiles, the follower-following network, and content (koos, rekoos, comments, likes, and mentions) posted on the platform, from Koo's public API that serves the Koo platform. Data collection started on 26 February 2021 and lasted until 11 March 2021. Table I summarises statistics of our proposed dataset.

Koo's stark resemblance to Twitter also invites a comparison between user behaviour on the two platforms. We created a dataset of Koo and Twitter user IDs that correspond to the same entity. This dataset can be used for automated identity resolution tasks and cross-platform analysis of Koo and Twitter.

We make our dataset public in adherence to FAIR principles as described in Section III-C.

Entity	Count
User profiles	4,061,735
Follower-following relationships	163,117,465
Koos	7,339,684
Rekoos	2,828,158
Rekoo with Comments	413,955
Comments	4,793,492

TABLE I: Statistics of the dataset we collected from Koo.

1) Koo Data Collection: We collected and analyzed data about the general discourse on the platform and the indulgent users, without restricting ourselves to certain trends, hashtags, or topics. Analyzing close-to-complete networks enables a significant understanding of the entire platform [24]. We used the follower-following network to discover and collect users recursively using the snowball methodology, essentially traversing the network graph in a breadth-first manner, similar to Kwak, et al. [2] and Zannettou, et al. [16]. We seeded the search with Koo's official language accounts (Table II) which are the first accounts shown to a new user, and the list of popular accounts that Koo recommends users to follow.¹⁰ These were chosen because of their large number of followers. We then collected the (previously undiscovered) followers and followees of these accounts and repeated the process for their follower-followee network. This approach is bound to leave out singletons and isolated communities; however, we hypothesize that they would not constitute a large share of the active users on the platform. We preferred this over enumerating and querying the API for all possible user IDs because of the significantly larger number of API requests the latter entails, which may have overloaded Koo's servers. We then used the user IDs to collect the users' profiles and the content (koos, rekoos, comments, likes, and mentions) generated by them, by querying the appropriate endpoints.

Language	Koo Handle	# of Followers	# of Following
Hindi	@कू_हिंदी	1,876,451	1
Kannada	@ಕೊ_ತಂಡ	710,509	1
Telugu	@కూ_కుటుంబం	259,767	1
Marathi	@कू_मराठी	67,383	0
Tamil	@தமிழ்கூ	35,698	0
Bangla	@কু_বাংঁলা	13,035	0
English	@kooenglish	316	0

TABLE II: Koo's official language accounts. Koo's Hindi account has the highest number of followers. Koo handles are written in the specific languages.

¹⁰https://www.kooapp.com/people

2) Twitter Data Collection: To create the Koo-Twitter user ID dataset, we considered two sets of users: those with verified Koo accounts and those who had listed their Twitter handles on their Koo profile. For the verified users, we manually curated corresponding Twitter handles. We queried Twitter's Users Search API for the users' names and found probable matches, manually annotating each one. 1,030 verified Koo handles were distributed equally among 10 annotators, who were asked to verify the corresponding Twitter handles of the users. According to a set of guidelines prepared by the authors, the annotators looked for similarity in profile pictures, profile information, and the content the users posted on the two platforms. All the annotations were verified by two annotators and ambiguous cases were dropped from the dataset. Out of the 1.030 verified Koo users in our dataset, we found matching Twitter profiles for 872 of them, 499 of which are verified on Twitter as well.

For the second case, we used Twitter handles provided by the users themselves. We removed duplicate handles and queried Twitter's Users Show API to eliminate invalid usernames. In all, we make public 38,711 Koo and Twitter user IDs that correspond to the same entity.

C. Adherence to FAIR Dataset Principles

The gathered data consists of publicly available information about a social network, gathering and examining which would provide significant insights into the platform's characteristics. Our dataset also conforms to the FAIR principles. In particular, the dataset is "findable", as it is shared publicly.¹¹ This dataset is also "accessible", given the format used (CSV) is popular for data transfer and storage. This file format also makes the data "interoperable", given that most programming languages and softwares have libraries to process CSV files. Finally, the dataset is "reusable", as the included README file explains the data files in detail. The data was collected through public API endpoints of Koo, adhering to their privacy policy.¹² The data we collected was stored in a central server with restricted access and firewall protection. All experiments shown in this paper were performed on this dataset.

IV. RQ1 : USER CHARACTERISTICS AND DEMOGRAPHICS

We analyze the demographics of users on Koo using their profile information. It should be noted that such user-entered information might not always be accurate and should be dealt with cautiously.

A. User Onboarding

The platform is reported to host 4.7 million users at the time of writing. Our dataset has approximately 4 million users, of which, 1.9 million joined Koo in the first two months of 2021 alone. Figure 1 shows the user creation timeline across various languages. We find that Koo was predominantly used by Kannada users in its initial stages, presumably because it was launched in Bangalore where Kannada is the vernacular dialect. Surges in influx of users can be seen in August 2020 - around

¹¹ https://precog.iiit.ac.in/resources.html

¹²https://www.kooapp.com/privacy. Accessed 12 March 2021.



Fig. 1: Weekly user creation timeline with language distribution, for 10 most popular languages. We observe peaks in users joining around August 2020, when Koo won the *Aatmanirbhar* Bharat App Innovation Challenge, and 10 February 2021, when MeitY tweeted about Koo. Kannada language was popular during the initial stages of Koo, following which Hindi language took over from August 2020. Many English users started joining from February 2021. The inset graph represents the lower frequency language users (languages except for Hindi and English).

the time of the *Aatmanirbhar* Bharat App Innovation Challenge Award. The user base expanded to other Indian languages, with Hindi users joining the platform in large numbers. February 2021 saw a huge spike in users with almost 200,000 of the users signing up on the days around 10 February 2021, just after the Ministry of Electronics and Information Technology, Government of India tweeted about the Koo app. As the debate of #KooVsTwitter trended on Twitter, many English users joined the Koo platform. The inset graph of Figure 1 shows the distribution of languages other than Hindi and English highlighting the prominence of Kannada, Telugu, and Tamil amongst Indian languages.

B. Gender Distribution

Of the 18.1% of the users who specified their gender on their profile, 92.1% identify as male (699,083), with only 7.5% users identifying as female (58,996) and 0.36% as others (3,236). However, Female users are more active, in terms of the number of average likes (103.6) and average rekoos (21.7) they do, as compared to other genders. Figure 2a also shows that Female have more followers (632.9) on average as compared to male users with an average of 117.0 followers and users identifying with the other category with an average of 283.45 followers. Male users, on average, produce fewer koos (5.9) and follow fewer people on average (84.8) than the other categories, as visible in Figures 2a and 2b. Figure 2c shows that the median age of is similar for all gender categories at 28 years. More male users identify themselves as single (12.0%), more female users as married (7.6%) and more users of the other category as divorced (3.0%), as shown in Figure 2d.

C. Language and Location

Koo allows users to choose their location from a list of Indian cities. In the 75,091 user profiles with location information, Bengaluru appears most frequently with 14,469 users, presumably because of the platform being headquartered in Bengaluru and being initially available in Kannada. Bengaluru also appears as a hint text over the location field on both the Android and iOS mobile applications. Mysore, the second most popular location, appears in only 1,469 profiles.

Despite Kannada being the first language on the platform, it is not the most popular, as it is outnumbered by Hindi, which was the language for 44.2% and 51.2% of the user base and total posts, respectively. Hindi was followed by English, which constituted to 23.8 % of the users and 25.9% of content (see Table III). The popularity of Hindi over English demonstrates a degree of success of the platform in promoting discourse in Indian languages. The distribution of users across languages closely mirrors the number of followers of Koo's official language accounts, apart from English (see Table II). This may be because posts from the language account corresponding to the user's language appear first in a user's feed.

D. Bio and Professional details

Koo allows the users to add and edit their profile bio. Figure 3 shows the word cloud of the user profile bios on Koo. We see the occurrence of words pertaining to state and national identities like "Marathi", "Tamil", "Bengali", "Indian", "Bharat" etc. For the 584,352 users who mentioned their work title, "student", mentioned 61,778 times, is the most common occurrence. "MBA" is the most common qualification with 4,988 mentions out of the 118,474 profiles that specify their work title.

V. RQ2 : CONTENT ANALYSIS

A. User-Generated Content

We have a total of 7,339,684 koos in our dataset. Figure 4 shows the post generation timeline of Koo with the language distribution. This displays a very similar behaviour to the



Fig. 2: Distribution of the user meta-data across genders. 18.1% of total users specified their gender, N = 761, 315, of which 92.1% are male, 7.55% are female and 0.362% belong to the others category. (a) The box plot corresponds to the 25th, 50th and 75th quartile. The iOS App Store policy mentions a minimum age of 12. Hence, we discard all ages below 12 (1.2%). Median age of 28 is observed across all users. (b) Female users have much higher average followers as compared to the other categories (c) Female users are more active in terms of likes and rekoos as compared to the other categories. (d) 12% of male users are single which is much higher than the other categories, while higher proportion of other category users are divorced.

Language	Number of Users	Percentage of Users	Number of Posts	Percentage of Posts
Hindi	1,795,411	44.2030	3,755,829	51.1715
English	968,271	23.8388	1,907,993	25.9955
Kannada	711,049	17.5060	818,679	11.1541
Telugu	259,171	6.3807	359,874	4.9031
Marathi	183,073	4.5072	242,803	3.3080
Gujarati	64,829	1.5960	126,853	1.7283
Tamil	48,285	1.1887	77,981	1.0624
Bangla	31,211	0.7684	49,172	0.6699
Malayalam	318	0.0078	257	0.0035
Assamese	46	0.0011	113	0.0015
Punjabi	43	0.0010	43	0.0005
Oriya	27	0.0006	87	0.0011

TABLE III: Language distribution on the platform. Hindi is by far the most popular language, for both user profiles as well as posting contributing to 44.20% of the user profile languages and 51.17% of the posts.



Fig. 3: Wordcloud for user bios. Presence of words pertaining to state and national identities such as "Tamil", "Marathi", "Gujarati" and "Indian" can be seen.

user creation timeline. A major spike in posting is observed around 10 February 2021, especially in English koos, when the Government of India's Ministry of Electronics and Information Technology posted a tweet promoting Koo, and many prominent government figures started joining the platform. The Kannada and Hindi language posts follow the same pattern as the user joining distribution. Similar to other social networks [25], the posting activity on Koo gradually increases throughout the day, attaining its peak in the evening at around 2100 hours IST.

B. Media Distribution

Koo allows for multiple types of media such as images, videos, GIFs, and text to be shared. Being a microblogging platform, more than 50% of the content on the platform appears as text, followed by other media. Most languages follow the same general trend, with a stark exception of Kannada, showing 40% of the content being shared through GIFs.

C. Hashtag Analysis

Hashtags in social media have become a way for users to build communities around topics and promote opinions. We extract the top occurring hashtags from the user posts and plot them as a wordcloud in Figure 5. We see hashtags like "kooforindia", "bantwitter", and "koovstwitter", which project a sentiment of competition between Twitter and Koo, and



Fig. 4: Weekly post creation timelines with language distribution, for top 10 used languages. A peak around 10 February 2021 is observed, when MeitY promoted Koo on Twitter. Posts were majorly in Kannada language during the initial days of Koo, following which Hindi posts took over from August 2020. Posts in English language spiked around February 2021. The inset graph represents the lower frequency language posts (languages, except for English and Hindi).

promote the Koo platform. Hashtags like "indiawithmodi", "atmanirbharbharat", "modi", "modistrikesback", and "bjp", which are associated with the Bharatiya Janata Party (BJP) are also present. Figure 6 shows a network of the 100 most frequently occurring hashtags, with the edges indicating two hashtags that occur in the same post. Although the graph has a low modularity score (0.329), hashtags related to one topic fall in the same category when the graph is clustered based on modularity [26].



Fig. 5: Wordcloud for top 25 hashtags in posts. Hashtags like #koo, #Koovstwitter, and #bantwitter show the competitive sentiment between the two platforms.

D. N-gram Analysis

In order to get an insight into the popular conversations going on the platform, we plot the top occurring uni-grams and bi-grams in the content of the posts (Figure 7). We observe an overwhelming number of Hindi n-grams. Through both uni-grams and bi-grams, we observe the existence of Hinducentric words like जयश्रीराम ("jaishreeram") and राम-राम ("ram-ram") that allude to "Lord Ram", a major deity in Hinduism. There is also a mention of many Indian religious leaders like रामपाल-जी ("rampal-ji"), संत-श्री ("sant-shri"), and जी-महाराज ("ji-maharaj"). This indicates that some of the discussions on Koo are around religion and that people may be using slogans like जयश्रीराम ("jaishreeram") as a symbol of their religious faith.



Fig. 6: A network of the 100 most frequently occurring hashtags, where an edge represents co-occurrence of hashtags in a post. Colors indicate clustering based on modularity. The entire graph has a modularity score of 0.329.

E. Mentions and Likes

The number of mentions and likes of a particular user are useful indicators to study a user's engagement and popularity on a platform [24]. Table IV shows the top 10 most mentioned users on the platform, of which "republic", which is an Indian news channel comes at the top with 16,041 mentions. Notably, Republic also has an editorial partnership with Koo.¹³ Ravi Shankar Prasad¹⁴ is the next most mentioned user with 12,991 mentions. Many prominent ministers and political figures like

¹³https://bit.ly/35nZFPJ Accessed 12 March 2021.

¹⁴Minister of Law and Justice, Electronics and Information Technology and Communications, Government of India.



Fig. 7: Word Clouds for user posts. Both unigrams and bigrams show substantial Hindu religion centric content.

Piyush Goyal¹⁵ and Sambit Patra¹⁶ are also present in the list of top mentioned users. Table V shows the top 10 users with the most liked posts on the platform. Ravi Shankar Prasad is the most liked user with 435,752 likes. Both English and Hindi pages for Republic News channel are present in the top 10 list.

VI. RQ3 : KOO'S USER NETWORK AND COMMUNITIES

We analyzed Koo's following network, looking for characteristics of the community forming as the young platform grows. The network has a noticeably high local clustering coefficient of 0.561, that represents how well connected the neighbourhood of a vertex is. This indicates a strong modular structure in the network, presumably due to Koo only catering to audiences from a single country. In contrast, Twitter, which caters to worldwide audiences, only had an average local clustering coefficient of 0.072 during its early years in 2009 [27], indicating much weaker communities.

We see that around 90% of the users have less than a hundred followers. However, a few users with an extremely high number of followers (of the order of 10^6) skew the distribution. The distribution of the number of followees follows a similar pattern. Further, users with a high number of followers tend to only have a small number of followees and vice-versa, indicating the presence of influential and popular accounts that are followed by a majority of other users. This may, in part, be due to new users being shown a list of popular accounts to follow, on top of their feed.

Figure 8 shows the following network between verified users on Koo. We see distinct communities of users based on language, possibly because users interact more with others whose language they can understand. English-speaking users are more centrally placed in the network, with connections to both Hindi and Kannada speakers. Verified accounts in the two Indian languages, on the contrary, do not have many follower-followee relationships.

VII. DISCUSSION

Koo's tagline, "The Voices of India", captures the essence of the platform, i.e., support for Indian languages, large Indian

¹⁵Minister of Railways, Commerce & Industry, Consumer Affairs and Food & Public Distribution, Government of India.

Handle	Mentions	Handle	Likes
republic	16,041	ravishankarprasad	435,752
ravishankarprasad	12,991	piyushgoyal	395,674
kisanektamorcha	11,010	republic	357,745
रिपब्लिक_भारत	9,366	रिंपब्लिक_भारत	296,039
piyushgoyal	9,127	meghupdates	277,495
mayank	7,588	rinki	257,266
leledirect.com	7,045	sawatimehera	239,962
aprameya	6,390	chouhanshivraj	185,354
sambitpatra	5,742	narendramodiforyou	181,271
khushbookapoor	5,693	anandranganathan	168,518

TABLE IV: Users with most number of mentions. News channel Republic has the highest number of mentions (16,041), followed by Ravi Shankar Prasad (12,991). TABLE V: Users with most number of likes. Ravi Shankar Prasad is the most liked user on Koo with 435,752 likes, while Republic takes the third place.

user base, and homegrown development. The recent surge in popularity of the platform, presence of multilingual content, and linguistic communities makes the study of Koo interesting. We release the first-ever dataset of the platform and characterize it based on the users, content posted, and the network. We show the formation of tight communities based on language, as well as the massive popularity of Indian politicians, news media agencies and government organizations on the network. We note that female users are more active, despite being present in smaller numbers. The higher presence of Hindi than English on Koo indicates a degree of success in promoting discourse in Indian languages. Kannada, Tamil, Telugu and Marathi also constitute a considerable portion of the content and activity on the platform. We observe a Koo vs Twitter rhetoric with hashtags such as "#koovstwitter" and "#bantwitter" trending. Koo is still in its nascent stages and is being developed to include more Indian languages - Gujarati, Malayalam, Oriya, Punjabi, and Assamese. As Koo grows in size, it should build convenient mechanisms with which researchers can collect and work on their data, which can prove useful for research in social computing and Indian languages.



Fig. 8: Following graph of verified users on Koo. Edges are directed from the follower to the followee. Node size is proportional to in-degree. Names are only shown for a few prominent accounts. Singletons are not shown.

¹⁶Official spokesperson of the Bharatiya Janata Party in India.

VIII. LIMITATIONS AND FUTURE WORK

This paper performs an exploratory analysis of the characteristics of Koo, presenting a novel dataset and uncovering valuable insights. Though our dataset contains a substantial portion of the user-base, it leaves out singletons and isolated communities. For future work, this multilingual data can act as a corpus for research in Indian languages. A study of posting trends by the same users on Koo and Twitter could reveal if users use the two platforms for different purposes, or if their activity on one mirrors that on the other. Our dataset of corresponding Koo and Twitter user IDs makes this data collection convenient. An analysis of users who gain popularity on Koo or the presence of bot accounts could reveal more about the inter-user interactions on the platform.

ACKNOWLEDGEMENTS

The authors thank the annotators for their contribution toward creating the dataset. The authors would also like to thank Nidhi Goyal, Samiya Caur and Shivangi Singhal for their helpful reviews and comments on the initial draft versions.

REFERENCES

- Pardeep Mittal and Amandeep Kaur, "E-governance: A challenge for india," *International journal of advanced research in computer engineering & technology*, vol. 2, no. 3, 2013.
- [2] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, "What is twitter, a social network or a news media?," in *Proceedings of the* 19th international conference on World wide web, 2010, pp. 591–600.
- [3] VenkataSwamy Martha, Weizhong Zhao, and Xiaowei Xu, "A study on twitter user-follower network: A network based analysis," in *Proceedings* of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, New York, NY, USA, 2013, ASONAM '13, p. 1405–1409, Association for Computing Machinery.
- [4] Cody Buntain and Jennifer Golbeck, "Automatically identifying fake news in popular twitter threads," in 2017 IEEE International Conference on Smart Cloud (SmartCloud), 2017, pp. 208–215.
- [5] F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C. Lu, and N. Ramakrishnan, "Misinformation propagation in the age of twitter," *Computer*, vol. 47, no. 12, pp. 90–94, dec 2014.
- [6] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze, "Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate," *American Journal of Public Health*, vol. 108, no. 10, pp. 1378–1384, 2018, PMID: 30138075.
- [7] Zeerak Waseem and Dirk Hovy, "Hateful symbols or hateful people? predictive features for hate speech detection on twitter," in *Proceedings* of the NAACL student research workshop, 2016, pp. 88–93.
- [8] Hajime Watanabe, Mondher Bouazizi, and Tomoaki Ohtsuki, "Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection," *IEEE Access*, vol. 6, pp. 13825–13835, 2018.
- [9] Anupam Jamatia and Amitava Das, "Part-of-speech tagging system for indian social media text on twitter," in Social-India 2014, first workshop on language technologies for indian social media text, at the eleventh international conference on natural language processing (ICON-2014), 2014, pp. 21–28.
- [10] Sudha Shanker Prasad, Jitendra Kumar, Dinesh Kumar Prabhakar, and Sukomal Pal, "Sentiment classification: an approach for indian language tweets using decision tree," in *International Conference on Mining Intelligence and Knowledge Exploration.* Springer, 2015, pp. 656–663.
- [11] Saifuddin Ahmed, Kokil Jaidka, and Jaeho Cho, "The 2014 indian elections on twitter: A comparison of campaign strategies of political parties," *Telematics and Informatics*, vol. 33, no. 4, pp. 1071–1087, 2016.

- [12] C Deep Prakash and Adrija Majumdar, "Analyzing the role of national culture on content creation and user engagement on twitter: The case of indian premier league cricket franchises," *International Journal of Information Management*, vol. 57, pp. 102268, 2021.
- [13] Gang Wang, Bolun Wang, Tianyi Wang, Ana Nika, Haitao Zheng, and Ben Y. Zhao, "Whispers in the dark: Analysis of an anonymous social network," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, New York, NY, USA, 2014, IMC '14, p. 137–150, Association for Computing Machinery.
- [14] Bahiyah Omar and Wang Dequan, "Watch, share or create: The influence of personality traits and user motivation on tiktok mobile video usage," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, pp. 121, 03 2020.
- [15] William A. Hamilton, Oliver Garretson, and Andruid Kerne, "Streaming on twitch: Fostering participatory communities of play within live mixed media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2014, CHI '14, p. 1315–1324, Association for Computing Machinery.
- [16] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn, "What is gab: A bastion of free speech or an alt-right echo chamber," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 1007–1014.
- [17] Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto, "Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system," in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2018, pp. 515–522.
- [18] Avinash Prabhu, Dipanwita Guhathakurta, Jivitesh Jain, Mallika Subramanian, Manvith Reddy, Shradha Sehgal, Tanvi Karandikar, Amogh Gulati, Udit Arora, Rajiv Ratn Shah, Ponnurangam Kumaraguru, et al., "Capitol (pat) riots: A comparative study of twitter and parler," *arXiv* preprint arXiv:2101.06914, 2021.
- [19] IV Kozitsin, AG Chkhartishvili, AM Marchenko, DO Norkin, SD Osipov, IA Uteshev, VL Goiko, RV Palkin, and MG Myagkov, "Modeling political preferences of russian users exemplified by the social network vkontakte," *Mathematical Models and Computer Simulations*, vol. 12, pp. 185–194, 2020.
- [20] Qi Gao, Fabian Abel, Geert-Jan Houben, and Yong Yu, "A comparative study of users' microblogging behavior on sina weibo and twitter," in *International Conference on User Modeling, Adaptation, and Personalization.* Springer, 2012, pp. 88–101.
- [21] Scott A Hale, "Net increase? cross-lingual linking in the blogosphere," *Journal of Computer-Mediated Communication*, vol. 17, no. 2, pp. 135– 151, 2012.
- [22] Scott A Hale, "User reviews and language: how language influences ratings," in *Proceedings of the 2016 CHI Conference Extended Abstracts* on Human Factors in Computing Systems, 2016, pp. 1208–1214.
- [23] Pushkal Agarwal, Kiran Garimella, Sagar Joglekar, Nishanth Sastry, and Gareth Tyson, "Characterising user content on a multi-lingual social network," in *Proceedings of the International AAAI Conference on Web* and Social Media, 2020, vol. 14, pp. 2–11.
- [24] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi, "Measuring user influence in twitter: The million follower fallacy," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1, May 2010.
- [25] Hai Liang and King-wa Fu, "Testing propositions derived from twitter studies: Generalization and replication in computational social science," *PLOS ONE*, vol. 10, no. 8, pp. 1–14, 08 2015.
- [26] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, "Fast unfolding of communities in large networks," *Journal* of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, pp. P10008, Oct. 2008.
- [27] VenkataSwamy Martha, Weizhong Zhao, and Xiaowei Xu, "A study on twitter user-follower network: A network based analysis," in *Proceedings* of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, New York, NY, USA, 2013, ASONAM '13, p. 1405–1409, Association for Computing Machinery.