2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Community Matters more than Anonymity: Analysis of User Interactions on the Quora Q&A Platform

Ehsan ul Haq*, Tristan Braud*, and Pan Hui*‡
*Hong Kong University of Science and Technology, Hong Kong SAR
‡University of Helsinki, Helsinki, Finland
Email: euhaq@connect.ust.hk braudt@ust.hk panhui@cse.ust.hk

***Abstract*—Question-and-answer (Q&A) websites are one of the latest evolutions in crowdsourced knowledge aggregation. Q&A websites provide more diverse opinions, as they involve the entire community. Quora made its reputation out of enhancing the traditional Q&A model with popular aspects of social media and incites its users to provide their names, locations, and references. This model allows higher quality control – including anonymous content, but more importantly, it leads users to form communities based on other criteria (e.g. profession, city) than similar interests. In this paper, we study the interactions among Quorans to unveil how such communities emerge. We perform both quantitative and qualitative analysis on the user-generated content and relate this content to social and demographic features. We show that being anonymous significantly affects the answers' length and subjectivity. On the other hand, most of the user interactions relate to their geographic locations.**

***Index Terms*—Anonymity, Demographics, Q&A Community**

I. Introduction

Quora[1] is a Q&A platform that integrates elements of social networks to the traditional Q&A model [1]. In parallel to these elements, Quora users can choose to hide their identities when posting. This anonymous model ensures a right balance between content moderation (only registered users can post anonymously) and the freedom of speech brought by anonymity (for sensitive or personal topics).

In Quora, each page is the result of the collective work of the community. When a user asks a question, it is automatically assigned to one or more topics by bots and can be refined by other community members. Similarly, the community can modify the question to make it clearer or merge it with another page. Apart from sharing information, Quorans can follow each other and the topics of their interests. Even if the exact ranking algorithm used to rank information is unknown, we know that it relies heavily on the aforementioned social features, including previous posting history, user ranking and popularity[2]. Quora offers a unique opportunity to study the interactions between users, especially personal information, such as demographics, permits to dissect the emergence of informal sub-communities in a system that pushes users to act as a whole united group. We theorize that these sub-communities are strongly related to the geographic location of Quorans, their knowledge of the English language and their cultural backgrounds. We compare these groups to both the anonymous and non-anonymous set of questions and answers.

Our contribution is twofold:

- Evaluation of the impact of anonymity on the tone and language employed, as well as the participation in topics.
- Interactions analysis between users of different countries, the impact of language knowledge and the main interests.

Earlier literature suggests that anonymity is used for self-expression and to avoid shame or social pressure [2]–[4]. Such use cases make anonymity a natural choice while sharing or asking personal experience as confirmed by the use of pronouns in Quora's questions [5]. Online participation is also highly linked with demographics and links between users [6]–[9]. In this paper, we further extend the study of anonymity and demographics; both are at interplay at Quora. We first focus on the effect of anonymity on the linguistic styles - length, polarity, and subjectivity of answers. Such factors play a critical role in the participation and appreciation of the content [10]. We then extend our study of user interaction to demographics and analyze its impact on topics participation.

We show that on Quora, anonymity has no effect on sentiments, but affects the linguistic style of the answers and is mainly used to talk about sensitive topics. Our study of demographics unveils the strong influence of user location on their interactions, as well as their center of interests and overall participation. Such analysis will help in better modelling user participation and ranking information.

II. Related Works

There are few studies directly targeted at Quora. A detailed study [11] discusses the relationships between different entities and their relevance for content recommendation. Another study [10] shows that anonymity leads to longer answers but has no effect on votes, views, and the overall politeness of answers. Matthew et al. [5] look at the linguistic styles of questions and shows that there is no significant difference between anonymous and non anonymous questions. Studies on other platforms show different relationships between anonymity and community behavior. Whisper and 4chan are 2 completely

[1]https://quora.com

[2]https://www.quora.com/How-does-the-ranking-of-answers-on-Quora-work

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00

anonymous social networks. A study [12] on *Whisper* shows that anonymous social interaction is short term. Besides, [13] compares content from Twitter and Whisper (anonymous) and display the different psycho-lingual features on both platforms. Bernstein et al. [2] show that anonymity induces rough and mob behavior on 4chan, but also highlights several positive effects on user participation. Kilner et al. [14] also show that anonymity increases the frequency of anti-social behavior among users. Cheng et al. [15] suggest that user' mood is one of the catalysts for trolling behavior. Another study [16] shows that group behavior becomes prominent for individual users in depersonalized groups. Omernick et al. [4] find that anonymity increases the curiosity of users. Finally, sometimes users can also opt for anonymity to escape governmental censorship [3].

In term of demographics, the cultural dimensions of Hofstede [17] show that the cultural background affects user behavior in a non-organizational context. Nistor et al. [18] shows the effects of New Media on inter-cultural interactions and conflicts. Studies on Facebook show that cultural values impact users interaction [6], [7], while the digital divide [19], also impacts the participation of some categories of users. Brailovskai et al. [20] exhibit positive relationship between one's self-presentation and social interactions for Facebook users from different countries. Another study [21] on Yahoo! Answers finds that the use of different languages across the platform significantly affects the quality of the content. Finally, demographic differences can also be attributed to the author's writing style [9]. We relate these theories to hypothesize our arguments on user interaction and anonymity. In this study, we identify certain interaction behavior based on demographics and extend the study of anonymity to relate it to stylometric variables such as subjectivity, polarity, and lexical diversity.

III. Dataset

As Quora does not provide any API, we rely on web crawling to collect data. We use `Chrome` web browser with the `Selenium` library in Python and parse the HTML with `BeautifulSoup`. We start with a single arbitrary topic (*Philosophy*) and recover the related questions. Each question leads to the user's profile, list of complementary topics, and answers. We collect Question, Answer, and User information for users posting answers or questions. We retrieved 1,645,845 questions linked to 146,617 topics. During this period, the number of questions with a complete log history is 680,802, and we extracted 1,411,397 answers for only 171,291 questions. We also retrieved 430,460 user profiles, 10,116 of which were deleted. Overall, 30% of questions and 3% of answers are from anonymous users, about 50% of users in our dataset have provided location information.

IV. Observations

We observe user interaction based on three aspects: anonymity, communities formed through Quora's social mechanics, and demographic communities.

Anonymity: We consider anonymous answers as being part of a separate community and focus on the impact of anonymity

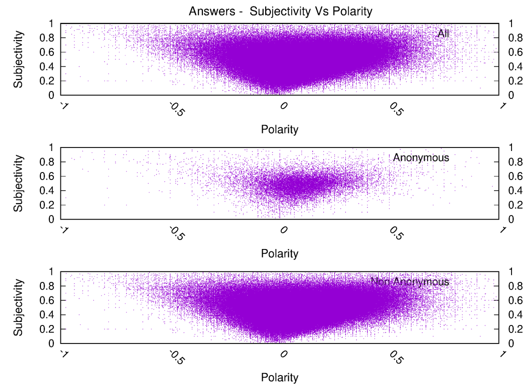


Fig. 1: Answers' subjectivity VS polarity. All (top), anonymous answers (middle), non-anonymous answers (bottom). Boudary polarity values lead to higher subjectivity.

on content. We first analyze the number of anonymous answers in each topic. We only consider topics with at least 20 answers, using similar approach by [5]. Topics with the highest anomymous answers to total answers include *stay-at-home spouses, thoughts, in love with best friend, secrets in life, strange stuff*, and *anus*, with ratios over 0.4. These topics are consistent with the definition of anonymous clusters by [5], having anonymity ratio $a_r > \mu_{a_r} + \sigma_{a_r}$. Where μ_{a_r} and σ_{a_r} are the mean and standard deviation of anonymity ratio across topics. Most of the topics presenting a high anonymity ratio are related to personal and private experience, which is concordant with the studies mentioned in the literature review section. This result confirms that anonymity allows people to talk about their personal experiences that they would not share otherwise. The ratio of anonymous questions is higher than anonymous answers (see Section III). Anonymity thus allows users to ask questions without fear of being judged or ridiculed.

Subjectivity and polarity are two primary measures in Sentiment Analysis. Polarity represents the user's sentiments, while subjectivity is the measure of the user's bias. We use both metrics to evaluate the average sentiments in anonymous and non-anonymous content. A study on Quora [10] showed that anonymity does not affect politeness. However, this study was based on a single topic and cherry-picked questions. We replicate these results on our large dataset and confirm that anonymity does not affect the sentiment of content on Quora, regardless of the topic. Figure 1 shows the relative subjectivity to polarity for all answers. We measure the polarity relative to the subjectivity for each answer and show them separately for all answers (top), anonymous (middle), and non-anonymous answers (bottom). This graph shows a consistent correlation between polarity and subjectivity. Anonymity does not affect the sentiments of content: most answers are centered in the middle, where both polarity and subjectivity are average. However, the higher the subjectivity, the more probable the answer is to have extreme polarity.

The ANOVA test shows a significant difference with $F = 148.6, p < 0.001$ in the length of answers and the lexical

TABLE I: Effect Size, Mean and Difference:

Parameter	Cohen's d	Anon	N-Anon	Diff.(%)
Length	0.321	114.8	82.62	39
Lexical Diversity	-0.243	0.79	0.83	-0.05
Subjectivity	-0.072	0.074	0.087	-0.15

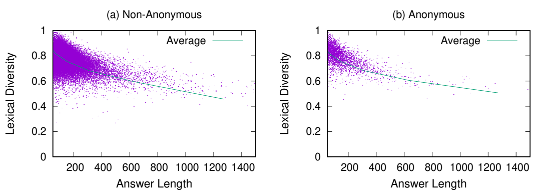


Fig. 2: Lexical Diversity vs Answer Length. Non-anonymous (Left), anonymous (Right). Answers with the same length have similar lexical diversity. Anonymous answers are longer with lower lexical diversity.

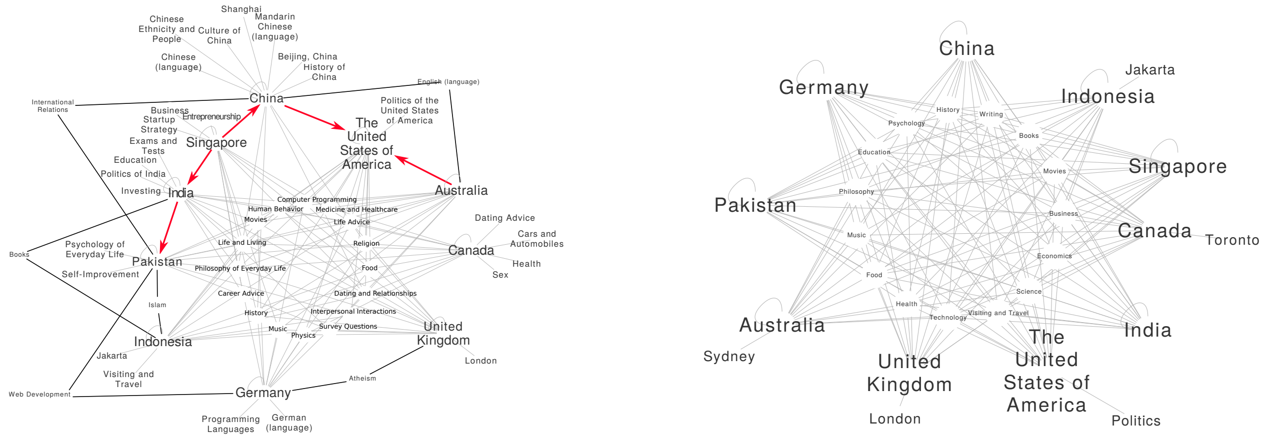
diversity $F = 59.8, p < 0.001$. We further report the effect size in Table I using Cohen's d test and difference measure [5]. The ANOVA test shows no statistical significance on sentiment polarity $F = 1.67, p > 0.05$ but the subjectivity of the answer is significantly different with $F = 23.20, p < 0.001$.

Lexical diversity is represented as the ratio of the unique word count to the total number of words in the text of a single answer. We measure the lexical diversity without removing any words from the answer's text. We show the results in Figure 2. Figure 2(a) shows the lexical diversity depending on answer length for non-anonymous answers while Figure 2(b) relates to anonymous answers. We draw three conclusions from this experiment: (1) the average lexical diversity for anonymous and non-anonymous answers of the same length is almost the same. (2) In general, anonymous answers are longer than non-anonymous answers. (3) for all answers, anonymous answers tend to have lower lexical diversity as compared to that of non-anonymous answers. This last phenomenon is due to the fact that lexical diversity is inversely proportional to the length of the text and anonymous answers are slightly longer on average. We also confirm the results of the previous study on anonymity on Quora [10], which stated that anonymous answers are usually longer than non-anonymous answers. A in-depth analysis of the vocabulary would permit us to fingerprint users and relate anonymous answers to registered users.

Official communities: Two kinds of communities emerge from Quora: first around users and the seocond around similar interests. We first consider communities formed around users that have more than 10,000 followers. 722 users in our dataset match this requirement and will be referred as "influencers". 43% of these users come from the US, and 36% are from India. In our dataset, most of the influencers belong to the following 3 categories: experts (engineer, researcher, attorney), artists (authors, poets) and politicians. Among these influencers, experts become popular through intensive participation on the website, while personalities tend to contribute more sparingly.

To understand how users group by interest, we analyze the most popular topics in our dataset and study how topics are related to each other. We first rank the most popular topics by *Interest* (topics most followed by users in our dataset) and *Participation* (the number of time a user answers to the question in specific topic). A user may follow a topic but not necessarily post content. We start with content followed by users. If a topic is repeated for n users, we increase the count for that topic by n. This gives the popular followed topics in our dataset. We then collect the topics. We add up the number of times users have answered in topics, and rank them according to the cumulative sum, resulting in a list of topics ranked by activity. Top ten followed topics include *Technology, Science, Books, Psychology, Movies, Education, Health, History, Music, Business* and top ten participated topics are *Philosophy of Every Day Life, Life Advice, Life and Dating, Dating and Relationship, Politics of the USA, India, Religion, Human Behavior, Computer Programming, Food.* All the topics followed are general topics such as Health, Movies, History, Science, and Business. On the other hand, users participate the most in topics related to human relationships. *India* and *USA* are two exceptions and can be explained by the proportion of people originating from these countries (see Section Demographics IV) and the prominent position of the USA in today's world's politics.

In a second experiment, we analyze how topics are interconnected. We collect all the topics followed by the users in our dataset and consider only those where users have written answers. We rank the topics by the number of answers posted. For each topic in this list, we then find the users that posted an answer and retrieve other topics these users have participated to. Each time a user participates in two topics, we increment the value of the liaison between those topics. As a result, we get a list of the topics ranked by the number of answers, along with the related topics, ranked by user participation. We only select the top 100 topics with the highest number of answers as our starting points. For each of these topics, we select the 100 most common connected topics. The resulting network graph is composed of 757 nodes – the topics – interconnected over 15,652 edges, of distance $-log(1/x)$, x being the number of interconnections between the two topics. Each node has, on average, 11.124 neighbors, with a characteristic path length of 2.677. To extract the main clusters out of this graph, we apply a single linkage hierarchical clustering algorithm and show the main clusters in Figure 4. For readability purpose, we only kept the 50 closest topics in each cluster. The biggest cluster is centered around the two topics *Life Advise* and *Life and Living*. From these topics emerge two other nested networks, one related to *Religion* and the other to *Relationships and dating*. The second biggest cluster revolves around *USA* and its politics. A branch of this cluster goes to the topics *International Relationships* and *China*. These two clusters confirm the analysis from our first experiment: users are the most active in topics related to personal life and the USA. Other major clusters include various common topics *Computer Programming*, *Quora*, *Science*, *Animals*, *English*



(a) Topic participation per country - Clear differences depending on the location of users. Some topics are shared only by a few countries.

(b) Topics Followed by country - Most users follow the same topics regardless of their locations.

Fig. 3: Cross country interactions

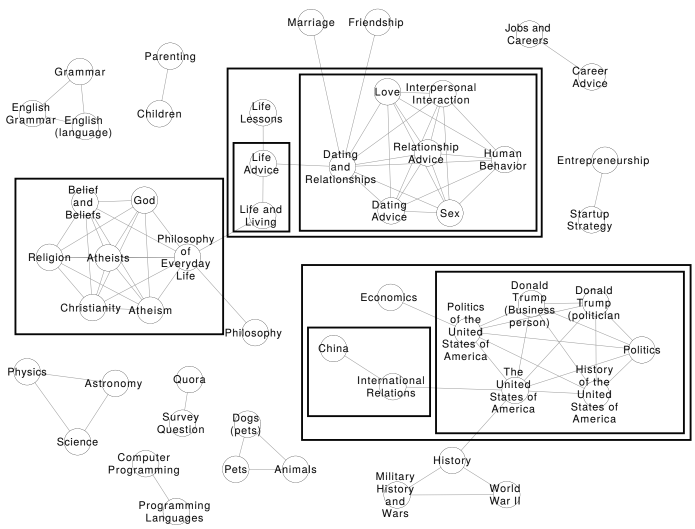


Fig. 4: Topic interaction clusters.

Language, *Parenting*, *Jobs*, and *Entrepreneurship*.

User Demographics: We analyze the most popular topics per country. Figure 3 represents the interactions between the 15 most popular topics for the 10 most active countries of our dataset, in terms of participation Figure 3(a) and interest Figure 3(b). The most active topics per country strongly depend on the demographics of users. All the countries are active in a unique subset of topics except The USA. We relate this phenomenon to two reasons: American users are ethnically more diverse and have higher topic similarity with other demographics. Specific topics show us the main preoccupations of people in a given country. For instance, users from China tend to participate exclusively in topics related to Asia. Indonesia and Pakistan only share the topic *Islam*. Figure 3(b) confirms the results from Section IV: most users follow general topics such as *Health, Music* or *Philosophy* and only a few countries show a clear identity. In this graph, the unique features for each country are mostly related to regional discussions or cities. Users from The UK, USA, and Canada do not have the topic of their country among the top 15 topics. These are also the general topics that Quora suggests to the user on sign up.

By digging deeper into the dataset, we can isolate more regional particularities. Topics can be the reflection of social and cultural openness towards different lifestyles and relationships. *Dating and Relationships* is commonly found in most countries from all the continents. However, topics like *Sexuality, homosexuality, Sex Advice* only appear in users from countries like The US, Germany, Canada, New Zealand, and the Netherlands. We also observe that users from Asian countries participate in topics like *Career Advice, Jobs and Careers, and Books*. Users from Nigeria, the United Arab Emirates and Hong Kong SAR show topics strongly related to their cultures and economic states. Topics related to the software industry and application development are more frequent for Nigerian users. UAE users are more interested in finance and economics, and HKSAR users heavily discuss topics related to cryptocurrencies and Chinese culture. Pakistani and Chinese users have International Relations in common. Another observation is the interest of users from one country in another country topics. We show it through red directed lines in Figure 3(a) Singapore users participate in *China, India*. While Indian users tend to participate in topic *Pakistan*. We represent in Figure 5 the heatmap of interactions between the 50 most active countries in our dataset. We consider as an interaction two users from different countries replying to the same question. As most of the users are from the USA and India, we represent the logarithm of these interactions to show interactions between less active countries. US users are more interaction than any other country. Users from the top countries also have significant interactions among each other. Users also tend to interact more together when they come from the same country, due to the presence of geographically-related topics.

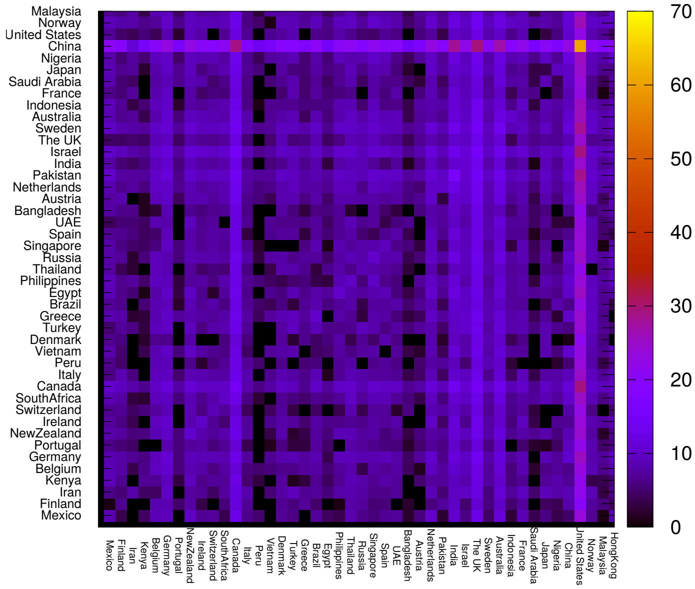


Fig. 5: Heat map of interactions between countries. People from the same country tend to interact more with each other than with other countries. Strongest interactions between native English countries and strong interactions between users of same countries.

We notice a more subtle continental and religious preference. For instance, Chinese Quorans interact slightly more with Taiwan, Hong Kong, Nepal, Malaysia, Japan, Indonesia and Pakistan, although we need to increase the size of our dataset to draw more robust conclusions.

V. Conclusion

In this paper, we demonstrated that anonymity does not affect the polarity, confirming the results of previous literature at a larger scale. Anonymous and non-anonymous answers significantly differ in terms of length, subjectivity and lexical diversity. We also showed that a higher subjectivity leads to more extreme polarity, potentially due to the self-experience discussed in the anonymous content. We showed that users form two different kind of communities: some communities are primarily based on the geographical locations and interests, while other communities are centered information and knowledge exchange. We categorized these communities as *Interest* (passive, centered around general culture) and *Participation* (active, often related to lifestyle). We finally highlighted how different countries share specific interests while still being invested in unique topics based on their socio-cultural values.

We believe this study is a step further in understanding user interaction within this knowledge-rich platform. In our next steps, we will consider the temporality to the data and study how interactions evolve over time. Future works also include linguistic fingerprinting of users to highlight new communities. Finally, we will gather a larger dataset to finely analyze the interaction among countries.

Acknowledgement

This research has been supported in part by project16214817 from the Research Grants Council of Hong Kong, and the 5GEAR project and FIT project from the Academy of Finland.

References

[1] Growmap, "Quora: Why you want to use this question and answer site." [Online]. Available: https://socialimplications.com/why-use-quora/

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

[3] R. Kang, S. Brown, and S. Kiesler, "Why do people seek anonymity on the internet?: informing policy and design," in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013.

[4] E. Omernick and S. O. Sood, "The impact of anonymity in online communities," in *Social Computing (SocialCom), 2013*. IEEE.

[5] B. Mathew, R. Dutt, S. K. Maity, P. Goyal, and A. Mukherjee, "Deep dive into anonymity: Large scale analysis of quora questions," in *International Conference on Social Informatics*. Springer, 2019.

[6] K. S. Al Omoush, S. G. Yaseen, and M. A. Alma'Aitah, "The impact of arab cultural values on online social networking: The case of facebook," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2387–2399, 2012.

[7] Q. You, D. García-García, M. Paluri, J. Luo, and J. Joo, "Cultural Diffusion and Trends in Facebook Photographs," in *Eleventh International AAAI Conference on Web and Social Media*, May 2017.

[8] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007, pp. 56–65.

[9] C. Bortolato, "Intertextual distance of function words as a tool to detect an author's gender: A corpus-based study on contemporary italian literature." *Glottometrics*, vol. 34, pp. 28–43, 2016.

[10] M. Paskuda and M. Lewkowicz, "Anonymous Quorans Are Still Quorans, Just Anonymous," in *Proceedings of the 7th International Conference on Communities and Technologies*, ser. C&T '15, pp. 9–18.

[11] G. Wang, K. Gill, M. Mohanlal, H. Zheng, and B. Y. Zhao, "Wisdom in the Social Crowd: An Analysis of Quora," in *Proceedings of the 22Nd International Conference on World Wide Web*. ACM, 2013.

[12] G. Wang, B. Wang, T. Wang, A. Nika, H. Zheng, and B. Y. Zhao, "Whispers in the Dark: Analysis of an Anonymous Social Network," in *2014 Conference on Internet Measurement Conference*.

[13] D. Correa, L. A. Silva, M. Mondal, F. Benevenuto, and K. P. Gummadi, "The many shades of anonymity: Characterizing anonymous social media content," in *Ninth International AAAI Conference on Web and Social Media*, 2015.

[14] P. G. Kilner and C. M. Hoadley, "Anonymity Options and Professional Participation in an Online Community of Practice," in *Proceedings of th 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years!*

[15] J. Cheng, M. Bernstein, C. Danescu-Niculescu-Mizil, and J. Leskovec, "Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17.

[16] T. Postmes, R. Spears, and M. Lea, "Intergroup differentiation in computer-mediated communication: Effects of depersonalization." *Group Dynamics: Theory, Research, and Practice*, vol. 6, no. 1, 2002.

[17] G. H. Hofstede, *Culture's Consequences: International Differences in Work-related Values/cG. H. Hofstede*. sage, 1982.

[18] N. Nistor, A. Göğüş, and T. Lerche, "Educational technology acceptance across national and professional cultures: a european study," *Educational Technology Research and Development*, vol. 61, no. 4, 2013.

[19] P. Norris, *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge University Press, Sep. 2001.

[20] J. Brailovskaia and H.-W. Bierhoff, "Cross-cultural narcissism on Facebook: Relationship between self-presentation, social interaction and the open and covert narcissism on a social networking site in Germany and Russia," *Computers in Human Behavior*, vol. 55, no. Part A, 2016.

[21] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and yahoo answers: everyone knows something," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008.