Analyzing Hate Speech Toward Players from the MENA in League of Legends

Sercan Şengün
Northeastern University
Boston, Massachusetts, USA
s.sengun@northeastern.edu

Joni Salminen
Hamad Bin Khalifa University
Doha, Qatar
jsalminen@hbku.edu.qa

Soon-gyo Jung
Hamad Bin Khalifa University
Doha, Qatar
sjung@hbku.edu.qa

Peter Mawhorter
Wellesley College
Wellesley, Massachusetts, USA
pmawhort@wellesley.edu

Bernard J. Jansen
Hamad Bin Khalifa University
Doha, Qatar
bjansen@hbku.edu.qa

ABSTRACT

We analyze hate speech toward the MENA players as a form of toxic behavior in League of Legends in-game and forum chats. We find that this kind of toxicity: (1) is initiated by one or two players; (2) sparks from criticizing the skills of team members; (3) can be elevated by frustration with game elements and hardware; and (4) can turn into personal clashes. There is also non-toxic use of abusive language, which stresses the importance of context-aware analysis (i.e., interpreting.

KEYWORDS

Online toxicity; games; ethnicity; League of Legends; MENA

Countries from Which There Was At least One Player in Chat Log Dataset
Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Kuwait, Lebanon, Libya, Malta, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, and the United Arab Emirates.

Collection of Forum Data
We searched the official Riot Games LoL forums to identify threads that concern players from the MENA region or topics regarding the Middle Eastern cultures. 89 forum threads were identified from EUNE, EUW, and North American (NA) context. Although our game data did not include any insights from the NA region, discussions from that server also had the potential to understand the inception and propagation of hate speech, considering the large number of Middle East diaspora living in North America. We read these 89 forum threads and coded the content. We also took notice of the nationality or ethnicity of the players whenever it was self-expressed. This coding was done since our data does not specifically signify the nationality or ethnicity of the players themselves.

what is actually toxic). Finally, we find evidence that the type of toxicity varies by server location, advising gaming companies to consider the location of players when setting up policies to mitigate hate speech.

1 INTRODUCTION

In this study, we analyze the inception and propagation of hate speech in virtual environments as a special form of online toxic behavior. Although many other kinds of online toxicity exist (i.e., in the form of in-game behavior, etc.), we only approach it through the lens of racist or cultural hate speech against the users from the Middle East and North Africa (MENA) in the League of Legends (LoL) community. To examine this, we perform qualitative analysis on chat logs of 200 randomly picked matches (i.e., multi-player game sessions) where at least one player was from a MENA country by reading and coding the dialogue of the whole match. Additionally, we code 89 threads from official forums where discussions regarding players from the MENA region took place.

LoL provides a testbed for online behavioral research, as it has a vocal community both playing the game and consuming related content online (e.g. forums, YouTube) [4]. Overall, previous research on toxicity in games has focused on predicting and regulating online toxic chat behavior through the efforts of developers [4,6] and linguistic analysis of player communication [5]. Adinolf and Turkay [1] analyzed player perceptions and coping strategies with toxic behaviors in e-sports games. Additionally, Birk et al. [2] investigated the effects of social exclusion on play experience and hostile cognitions in games. However, culturally-specific toxic behavior in online games has not previously been studied in detail.

2 METHODOLOGY

Data Collection

For data collection, we collaborated with Riot Games, the developer of LoL, and received chat logs from their servers. Since there is no server specific to the MENA region, we received data from 2 separate servers: Europe Nordic & East Server (EUNE) and Europe West Server (EUW). Although there are no restrictions for the European Region LoL players about which server they can play in, players tend to select servers that are geographically closer to them to avoid lag issues and stay close to their communities.

From each server, we selected and anonymized chat data from 30K random matches that involved at least one player from 17 MENA countries (see sidebar) for a total of 60K matches that took place sometime during the last quarter of 2017 and the first quarter of 2018. The country check was done using player IPs and did not indicate ethnicity or nationality but only that those specific players logged into the game from an IP associated with a MENA country. Although VPNs would let players outside the region appear to be in it, we assume the use of such would be negligible since it would create lag issues that players would rather avoid. Additionally, since the

Table 1: TOP15 toxic words in the LOL chats. Numbers include both EUNE and EUW servers.

Toxic word	Count	Frequency
noob	13,669,004	26.2 %
fuck	7,546,548	14.5 %
stfu	3,810,928	7.3 %
idiot	3,438,006	6.6 %
noobs	2,705,776	5.2 %
fucking	2,503,765	4.8 %
suck	1,985,366	3.8 %
fuckk	1,980,044	3.8 %
retard	1,855,247	3.6 %
fking	1,722,399	3.3 %
shut up	1,519,692	2.9 %
useless	1,382,759	2.7 %
bitch	1,175,127	2.3 %
trash	735,918	1.4 %
fuking	649,124	1.2 %

Table 2: Self-expressed nationalities of the players in 89 forum threads.

Forum region	Distribution of nationalities (counts for server; percentages for total)
NA	Egypt (1), Jordan (1), Saudi Arabia (1), United Arab Emirates (3)
EUNE	Algeria (1), Bahrain (4), Egypt (15), Iran (3), Iraq (1), Jordan (3), Kuwait (7), Lebanon (22), Oman (4), Palestine (2), Qatar (2), Saudi Arabia (4), Syria (3), United Arab Emirates (79)
EUW	Bahrain (3), Egypt (5), Iran (1), Jordan (4), Kuwait (3), Morocco (2), Oman (3), Qatar (3), Saudi Arabia (6), Syria (1), Tunisia (1), United Arab Emirates (14)
Total	Algeria (0.5%), Bahrain (3.5%), Egypt (10.5%), Iran (2%), Iraq (0.5%), Jordan (4%), Kuwait (5%), Lebanon (11%), Morocco (1%), Oman (3.5%), Palestine (1%), Qatar (2.5%), Saudi Arabia (5.5%), Syria (2%), Tunisia (0.5%), United Arab Emirates (48%)
Note: players from the United Arab Emirates (48%) being the most vocal (or more likely to express their nationalities), followed by Lebanon (11%) and Egypt (10.5%).	

game is not banned in any of these countries, the players would not have any specific incentive to use an IP proxy service.

Moreover, we code 89 discussion threads from the official LoL forums where issues regarding MENA players and cultures were discussed. We refer to this dataset as "forum data" (see the sidebar on this page, and Tables 2 and 3 for descriptions of the forum data). For each server, we had around 1–2 million lines of chat. To analyze the data, we read a random sample of 5,000 messages in order to create a lexicon of hateful (n=48) words and their variations (see Table 1). We used this lexicon to identify the matches where hateful chats took place. One researcher qualitatively analyzed 100 matches from each server by reading and coding the chat logs for a total of 200 matches. These matches were randomly selected among those that had at least one hateful word in them. All the chat within those games was read, analyzed, and qualitatively coded. In the process, we created suitable categories to reflect the research aim of analyzing the inception and propagation of toxic hate speech in this environment. Our approach follows the open coding procedure of Strauss and Corbin [7].

Scope of the analysis

Here, we outline the chat and matchmaking options in LoL and how they affect our data. In terms of chat options, players can: (1) type in the chat box to send messages to everyone in the match, only to their teammates, or privately to one another—they can also ignore each other; (2) adjust push-to-talk options and participate in the voice chat where the options are similar to the text chat; and (3) report portions of text chat and other players' behaviors which usually initiates a long, non-transparent, and inconclusive process. It is important to note that some discussions might have been carried over to the voice chat that we did not have access to.

Matches in LoL can be played in normal or ranked modes (a crude comparison of these modes might be hobbyist vs. professional) and in multiple map types. Our random sample did not take these factors into consideration other than the match containing at least one player from the MENA.

3 EXPLORATIVE RESULTS FROM CHAT LOGS

We identified 1058 toxic segments from the chat logs of 200 matches. A toxic segment consists of a chat line that contains at least one toxic word as well as the related exchange that follows. First, we look at the toxicity count tendencies (segments/match): $\bar{x}$ = 5.305, median = 4, R = 30. See the histogram in Figure 1 where we eliminated one outlier match with a very high (59) toxicity count. Without this match the tendencies change to $\bar{x}$ = 5.035, median = 4, R = 12.

We also look at the number of players who are involved in the toxic chat. In the standard match data that we have, each match has 5 players on either side to a total of 10. The tendencies for the number of players who are involved in the toxic chat are: $\bar{x}$ = 1.685, median = 1, R = 2.5 (see Figure 2). These numbers show us that toxicity in in-game chat is mostly initiated and pursued by one or two players; most other players do not get involved.

Table 3: Racism and/or Islamophobia targeting the Middle East in 89 threads from the League of Legends official forums. There was more hate speech in EUW and NA than in the EUNE forum.

Forum region	Racism and/or Islamophobia
NA (%)*	22 (36%)
EUNE (%)*	15 (25%)
EUW (%)*	24 (39%)
Total (% to the total)	61 (21.6%)
* Percentage of the codes among servers	

Trolling or Hate Speech?

Frequently, it is possible to find comments that border between hate speech and trolling. Harrell et al. [3] define the distinctions and the commonalities between sarcasm/irony, trolling, and hate speech for the purposes of the representations of African cultures in video games.

In our context, an example would be: "[Riot Games] doesn't want an [A]rabic server, they are afraid it might go kaboom one day." This innuendo is in-line with further trolling where Arab players are associated with LoL characters that use bombs or "blow themselves up." Consequently, it is hard to differentiate if a comment represents a genuinely wrong perception of a culture, is trolling, or is intentionally abusive. Additionally, trolling might be adopted as a way of retaliation, bargaining, or diminishing tension (e.g., "I find it more hilarious if the [Z]iggs player was American. We bomb way more people... [J]ust sayin'.")
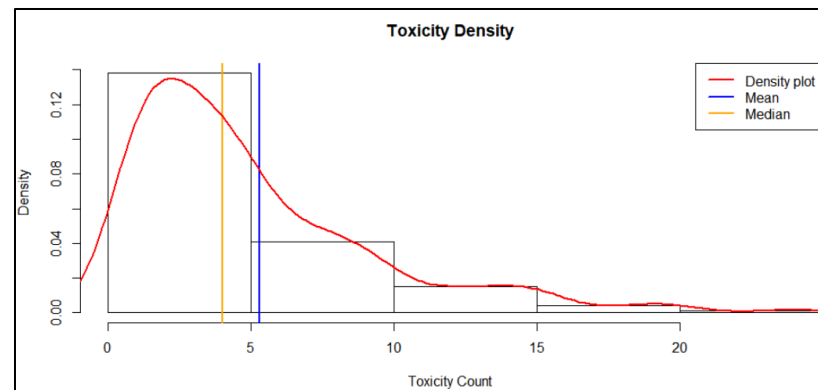

Figure 1: Measures of central tendencies in toxicity count per match
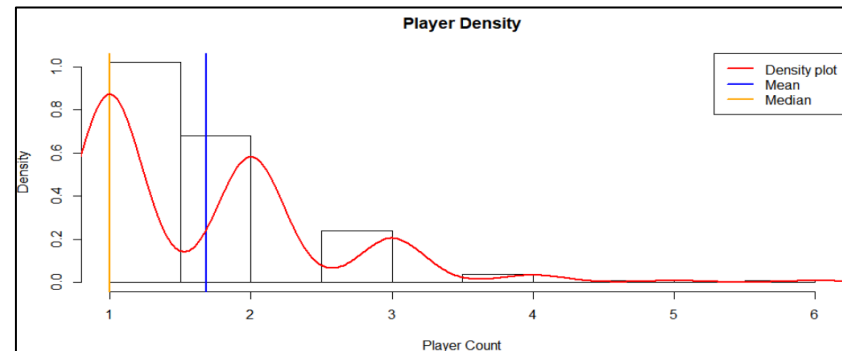

Figure 2: Measures of central tendencies in the number of players getting involved in toxic chat

According to the Pearson correlation coefficient, there is a moderate positive linear relationship between the toxicity count and player count (r = .617, p = 2.2e-16). This is unsurprising, as more players get involved in the discussion, the opportunity for toxicity gets amplified.

We checked the language used in the chat for these matches. 155 out of 200 matches were in English only (77.5%), 31 were Arabic and English mixed (15.5%), and 14 were a mix of English and a non-Arabic language (7%). Since LoL does not support Arabic characters, when we refer to Arabic use in the in-game chat we are actually referring to the use of *Arabish* or *Arabizi* [8] which are different names given to the practice of typing Arabic with a combination of Latin letters and numbers. Our analyses excluded chat typed in this way as well as other non-English chat. As a result, we found no significant correlation between the languages used in a match and the toxicity count.

Table 4: Examples of toxicity based on race, ethnicity, or culture from chat logs.

Targeted toxicity	Example
Skills or the incompetency of the Arab players	"Arabs... noobs," or "he is a fucking noob [A]rabian."
Referring to an Arab player as Ziggs (suicide bomber champion) although the player is not using that character	"Next time you are dead noob Ziggs [...] lol I troll vs Ziggs."
Referring to an Arab player with stereotypical Arabic names although that is probably not the name of the player	"sorry Ahmed, fffff, go kill yourself," or "mohamed come noooob."

Examples of Racism and/or Islamophobia in Forum data
"Aren't you all running around on the streets with AK-s and selling drugs and bombs to each other? How is that you are on the internet? :O"
"Are women allowed to join this [...] or women can't play [LoL] in your countries? [Do] you throw rocks [at] them if they are caught playing [...]?"
"Shouldn't you be blowing stuff up instead of playing LoL?"
"You ruined our country with your retarded desert culture now you want [to] ruin our servers too? Stay away please."
[In response to Arabic language support] "Because [you] could talk about plans [on] how to bomb Europe, and noone would understand it."
"You plan to go all JIHAD on the enemy team?"

4 TOXICITY CATEGORIES FROM CHAT LOGS

Additionally, we coded the 1058 toxic segments to understand their content. The results are given in Table 5 and each category is discussed in the following.

Non-toxic Use: Occasionally, toxic words are used within a non-toxic context as parts of "friendly quips." These uses usually emerge when players know each other from before. This use was significantly more visible in EUNE than in EUW. Some examples include: "save me babies [...] danny help me bitch," "I didn't see you noob sorry," and "yea :DDDD noob [...] you go top or bot."

Game Elements: Players regularly curse elements of the game with lag, latency, and ping issues deserving a special mention. The other game elements mentioned are champion design, NPCs (minions), and players' own equipment (e.g., mouse, video card, etc.).

Gaming Skills: The most important cause of toxic behavior begins with players judging each other's skills in playing LoL. In many cases, these provocations are left unanswered. When the other user responds to the accusations of poor play, the discussion quickly turns personal. (More analyses on this is presented below.) Compared to the "personal," this category is more focused on in-game elements and events and less on personal attributes or information. In this regard, they can be comparatively less toxic but are also predictors of more personal toxicity. An interesting observation is that these discussions typically take place among the team members. Inter-team toxic discussions on game skills are scarce, likely in part because by default chat is sent only to one's own team.

Personal: The personal discussions category can be an escalation of gaming skills discussions or start by other means. We use this coding when the context of discussions and the use of toxic language transcends the boundaries of game elements, game events, and gaming skills, and become personal. Again, it is important to note that these discussions are also predominantly among team members. We have found extremely little evidence of toxic personal discussions between teams. Table 4 and the sidebar on this page includes examples of racial toxicity. We analyze how the escalation of toxic chat can be predicted between the categories (see Table 6).

There is not a drastic difference between the servers for general toxic chat. In terms of hate speech on racial, ethnic, or cultural context, EUW leads significantly. there was no such discussion in our sample set within EUNE. Although this might be a result of the randomization or of EUNE community being more used to Arabic-speaking players or a sample size, this is in-line with our observations that racist or Islamophobic comments were less likely to appear in the EUNE forums.

5 DISCUSSION AND IMPLICATIONS

This study provides novel insights on in-game toxicity through the lens of culturally-specific hate speech. Based on reading and coding both the chat logs and forum messages, we note that simple dictionary-based techniques are insufficient to detect, understand, and moderate such toxicity. Rather, context-awareness is needed as the interpretation of hate is based on how the words are used and what the language norms within a team and/or forum context are. Even then, the distinction between trolling, cultural misconceptions, and purposeful abusive (toxic) use of

Table 6: Pearson correlation coefficients between the segment counts of toxicity categories (excluding the non-toxic use category as it does not correlate with any other categories)

	Game Elements	Gaming Skills	Personal
Game Elements	1	$r = .359$ $p = 1.837e-07$	$r = .316$ $p = 5.146e-06$
Gaming Skills		1	$r = .215$ $p = .0023$
Personal			1

Interpretation: There is a significant positive correlation between gaming skills and personal categories. This leads us to believe that trash talking one's own team's skills can escalate into personal insults. Even more interesting is the significant and moderate positive correlations between game elements with gaming skills and personal categories, which suggest that toxic talk about game elements may lead to toxic remarks about gaming skills or personal insults.

Implications of Findings

Overall, the prevalence of hate speech and racism can vary by geographic location of the servers/forum, which is a detail that gaming companies should recognize in the context of moderation.

One likely explanation to the variation is that a large share of Middle Eastern users in a server/forum results in natural lack of hate speech, possibly via mechanisms of self-moderation.

Thus, assigning players to culturally cohesive groups could mitigate race- or ethnicity-related toxicity.

language can be difficult to ascertain. This implication carries weight for researchers, game designers, and moderators–the importance of manual analysis is particularly topical, as the current mainstream of toxicity research tends to focus on automated techniques rather than in-depth analysis of toxicity. Future work on natural language processing can also benefit from addressing context-sensitive analyses techniques. Furthermore, the finding that a small number of instigators, typically one or two out of 10-players in a match, is responsible for the most toxicity in online game chats implies that online gaming companies should seek, isolate, and ban players that exhibit continuously toxic behavior.

Table 5: Categories of 200 online LoL matches involving at least one MENA player and at least one toxic keyword.

Category	Sub-category	EUNE server		EUW server	
		Count	%	Count	%
Non-toxic Use	Game Element	1	0.18	1	0.20
	Own Team	22	3.91	7	1.41
Game Elements	Lag	15	2.67	12	2.42
	Other	17	3.02	11	2.22
Gaming Skills	Other Team	23	4.09	43	8.67
	Own Team	171	**30.43**	175	**35.28**
	Self	32	5.69	42	8.47
Personal	Other Team	5	0.89	9	1.81
	Own Team	276	**49.11**	186	**37.5**
Racial, Ethnic, or Cultural		0	0	10	2.02
Total		562		496	

REFERENCES

[1] Sonam Adinolf and Selen Turkay. 2018. Toxic Behaviors in Esports Games: Player Perceptions and Coping Strategies. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, ACM, 365–372.

[2] Max V. Birk, Benjamin Buttlar, Jason T. Bowey, et al. 2016. The Effects of Social Exclusion on Play Experience and Hostile Cognitions in Digital Games. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 3007–3019.

[3] D. Fox Harrell, Sercan Şengün, and Danielle Olson. 2019. Africa and the avatar dream: Mapping the impacts of videogame representations of Africa. In K. Josephs and R. Risam, eds., *The Digital Black Atlantic*. University of Minnesota Press, Minneapolis.

[4] Yubo Kou and Bonnie Nardi. 2013. Regulating anti-social behavior on the Internet: The example of League of Legends. *Proceedings of iConference*, 616–622.

[5] Haewoon Kwak and Jeremy Blackburn. 2014. Linguistic analysis of toxic behavior in an online video game. *International Conference on Social Informatics*, Springer, 209–217.

[6] Brendan Maher. 2016. Can a video game company tame toxic behaviour? *Nature News* 531, 7596: 568.

[7] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. *Handbook of qualitative research*: 273–285.

[8] Mohammad Ali Yaghan. 2008. "Arabizi": A Contemporary Style of Arabic Slang. *Design Issues* 24, 2: 39–52.