Increased Harassment of Female Journalists of the Iranian Diaspora During the Iranian Protest of 2022

Abstract: Female journalists often experience considerable harassment targeted at their occupation, which is significantly intensified on social media platforms. Our study investigates this phenomenon within a political context, specifically focusing on female journalists from the Iranian diaspora. We examine a new harassment campaign targeting female journalists, researchers, and activists who reported on the 2022 Iranian protests. Our research reveals that female journalists experience more online harassment compared to non-Iranian journalists and male journalists from the Iranian diaspora. We also discovered that these individuals faced a higher rate of online abuse on Twitter, largely driven by bot accounts within the network.

Introduction

Motivation & Background:

On September 16th, an Iranian citizen, Mahsa Amini, died allegedly after suffering injuries at the Vozara Detention Center [1]. Amnesty International claimed that Mahsa Amini was initially arrested on 13th Sept 2022 by the morality police in Tehran for not abiding by the country's dress code. [1] Allegations claim that Amini was beaten by the morality police which resulted in her suffering from a coma before passing away. [1] The authorities maintained that Amini died from a heart attack, but her parents dispute this official verdict. [2] The protest began on 17th September 2022 at the funeral ceremony for Amini and soon spread across Iran. Protesters took to the streets chanting slogans against the Iranian regime and demanding action against the individuals involved. [2] The security forces retaliated by using excessive force against protesters, which resulted in the reported death of 479 people, including 68 children. [3] Additionally, the regime limited internet access in the country to prevent protest coordination, as well as to limit information about the government's physical abuse of the protesters.

During this period of political tension in the country, journalists, particularly female Iranian journalists, played a vital role in making information related to the protests available to the public. In fact, one can argue that female journalists were at the forefront of this movement. The story of Mahsa Amini's torture at the hands of the Iranian morality police was exposed by a female Iranian journalist, Nilufar Hamadi. [4] Hamadi had been following cases of female arrests in Iran for some time and when she heard the news of Amini's coma she visited her in the hospital.[4] Hamedi shared a picture of Amini's bruised body and her distressed parents which went viral both nationally and internationally. [5] The anger experienced by the Iranian audience online quickly morphed into on-ground protests against the government. Hamadi was instrumental in breaking the news and setting off a series of events that resulted in thousands of people protesting all around Iran. In addition to Hamedi, Yalda Moaiery covered the protest from the ground. Her photographs of the violence displayed by the Iranian forces quickly went viral and provided credible evidence for the violent suppression of protesters by the government. [6] Both Hamedi and Moaiery were arrested between 19th and 21st of September 2022. The security forces confiscated all their journalistic materials and equipment. According to the Coalition for Women Journalists (CFWIJ), 25 female journalists were arrested since the beginning of the arrests, of which 12 were arrested within 48 hours of Amini's death.[7]

While female journalists in Iran faced imprisonment, those in the diaspora experienced increased harassment on social media platforms when they reported on the protests.[8] Although female Iranian journalists and activists who were not present physically in Iran avoided detainment, many reported facing increased abuse and threats online.[8] The targeted individuals reported that they were subjected to a coordinated campaign aimed at undermining their credibility. This involved a Twitter army spreading Tweets urging the journalists' employers to dismiss them, accusing them of being agents, mouthpieces, and lobbyists for the Iranian government. [9] This campaign claimed that these individuals were not representative of the Iranian people and had been working against their interests covertly. The harassment seemed to target female journalists and activists who are quite respected in their field and have been known to critique the Iranian government in the past. This campaign of harassment seemed to have targeted a substantial number of female reporters/activists. This increased provocation resulted in BBC filing a concern with the UN in regard to harassment of the Persian service journalists. [10] Additionally, The Coalition For Women In Journalism issued multiple press releases bringing attention to this increased phenomenon.[11] It also seemed that more female activists and journalists were coming forward claiming to have experienced increased harassment as compared to male Iranian journalists and non-Iranian journalists who were also reporting on the same events.

The harassment and abuse faced by female journalists in Iran and in the diaspora highlight the challenges faced by women in the media industry. Despite increasing recognition of women's contributions to journalism, there are still significant barriers to their full participation and protection. It is essential to continue advocating for the rights and safety of female journalists and to promote gender equity in media organizations. In this investigation, we wish to further explore this particular phenomenon of online harassment campaign faced by female journalists of the Iranian diaspora. In this research, we investigate how female journalists, academics, and activists of the Iranian diaspora came to experience a targeted defamatory campaign that sought to attack their credibility. In this research, we will also compare the harassment experienced by female journalists/ activists of the Iranian diaspora and compare the metrics against the harassment of other groups such as male Iranian journalists and non-Iranian journalists to see if a particular group was harassed disproportionately compared to the other groups.

Literature on the Online Harassment of Journalists:

As posting journalistic work on social media has become a popular means for journalists to increase their following online, the digital space has established a new sphere where female journalists experience more abuse. [12] This is not only true for journalists but also for female influencers online. Women YouTubers have been observed to face a significantly more hostile environment than their male counterparts. [13] Similarly, on Twitter, violent anti-feminist hashtags are regularly used against female journalists to reinforce gender stereotypes and create a fertile ground for defamatory attacks. [14]

Harassment can be defined as the use of incivility when referring to someone that goes beyond mere impoliteness and is marked by profanity, insults, name-calling, and challenging credibility. [15] Recent research has shown that women journalists across diffident cultures and regions are similarly feeling increased harassment on their official social media accounts. According to recent findings by the

International Women's Media Foundation, at least 70% of female journalists in the U.S. face regular online harassment. [16] In another study by Demos, a United Kingdom-based think tank, "journalism is the only category in the study where women received more abuse than men, with female journalists receiving roughly three times as much abuse as their male counterparts". [17] Additionally, in a survey held by the International Women's Media Foundation in 2018, it was found that 2/3 of female journalists experience online intimidation in relation to their journalistic publications. [18] Similar treatment of female journalists has been observed across cultures and regions.

In the paper 'You Really Have to Have a Thick Skin': A Cross-cultural Perspective on How Online Harassment Influences Female Journalists", the authors interviewed and collected data on 75 female professional journalists working in Germany, India, Taiwan, the United Kingdom, and the United States. [19] The study found that all participants experienced increased harassment once activating their social accounts. The study further identified that the majority of the participants found some level of difficulty interacting with the comments online which caused discomfort in performing their journalistic routines online. [19]

In 2018 Reporters Without Borders published a report on the online harassment of journalists worldwide. They focused on the case of Maria Ressa, the founder and executive editor of the news website Rappler in the Philippines. The study found that the author began to experience significant harassment online, threats of rape, murder, and arrest on social media as a result of publishing a critique of the government. [20] Similarly, Rana Ayyub, a freelance journalist from India also received regular death threats online by Prime Minister Modi's trolls, Yoddhas, for her investigative reporting on Modi's rise to power. [21] We find evidence of such a phenomenon at the more aggregate level as well. According to an analysis of 70 million reader comments on the Guardian newspaper from 2006 – 2016, author Beck Gardner found that females experienced more harassment in 5/6 categories (life & style, world news, technology, film and sports) compared to their male counterparts. [22]

Social Cyber Security: As research into Information Operations has become popular due to increased coordinated inauthentic behavior from state and non-state actors on social media, an important development in the area of the study is the growth of the academic field known as Social Cyber Security. This discipline studies the particular mechanisms through which agents manipulate information online.[23] Social Cyber security is now recognized by the National Academies as a new scientific discipline.[24] The discipline has been used to investigate information maneuvers, network maneuvers, information diffusion, and motive identification in the context of political disinformation. [23] This methodology is also becoming increasingly integrated into other disinformation investigations, such as medical disinformation, particularly around Covid-19.[23] In our research, we will use the principles of social cybersecurity to track online conversations using a network science approach.

The BEND Framework:

Traditionally, research related to bot participation in Twitter conversations would give us information about the presence of bots in the network but no information about what the bots were trying to do, a valuable insight. Therefore, the discipline of social cybersecurity strives to understand both the motives and tactics employed by the agents. [23] It does so through employing BEND Framework, a model that was developed to "assist in the theoretical conceptualization of this problem by providing a taxonomy of

16 categories of maneuvers for conducting online influence".[25] Carley and Beskow explain that these 16 maneuver categories are divided into 2 main categories of network and narrative maneuvers. The authors explain that 'Narrative Maneuvers' concentrate on influencing the content of shared posts in the network, thus affecting what is being said and discussed; Network Maneuvers on the other hand seeks to impact the structure of the network in question, i.e how the communities are shaped and what actors have key roles in the network. [23]

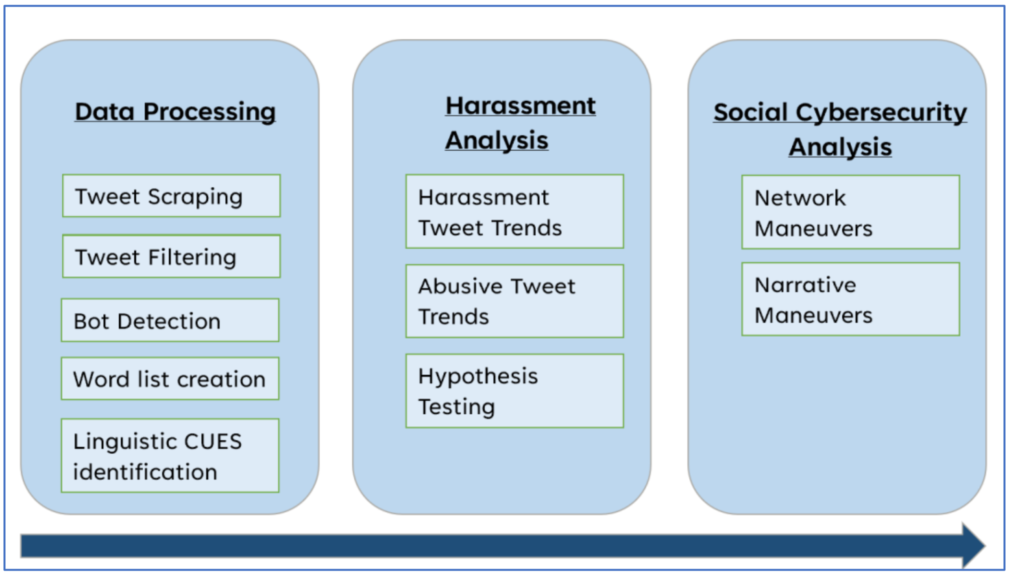
The table below lists the 16 maneuvers that can be applied to the information or the network itself and what they mean:

	Narrative Maneuver Knowledge Network Manipulation Things you can do by affecting what is being said	
Positive	Engage	Discussions that bring up related but relevant topic
	Explain	Discussions that provide detail on or elaborate the topic
	Excite	Discussions that bring joy/happiness/cheer/enthusiasm to group
	Enhance	Discussions that encourages the group to continue with the topic
Negative	Dismiss	Discussion about why the topic is not important
	Distort	Discussions that alter the main messages of the topic
	Dismay	about the topic which will bring worry/sadness to the group
	Distract	Discussions about the totally different topic and irrelevant

	Network Mauver Social Network Manipulation Things you can do by affecting who is talking/listing to whom	
Positive	Back	Actions that increase the importance of the opinion leader
	Build	Actions that create a group or the appearance of the group
	Bridge	Actions that build connections between the two groups
	Boost	Actions that grow the size of the group or make it appear as it has grown
Negative	Neutralize	Actions that limit the effectiveness of opinion leader
	Nuke	Actions that lead to a group being dismantled
	Narrow	Actions that lead to a group becoming sequestered from other groups
	Neglect	Actions that reduce the size of the group or make it appear that the group has grown smaller

Methodology

To conduct our analysis, we refer to previous papers written in social cyber security research for identifying necessary research methodology. Uyheng, Magenlinski, Villa-Cox, Sowa and Carly conducted a social cyber security analysis under the title, 'Interoperable pipelines for social cyber-security: assessing Twitter information operations during NATO Trident Juncture 2018', where they used social cyber security tools and methodologies to uncover bot propagated campaign on Twitter to discredit NATO's reputation online.[26] The paper executed the research by creating a data processing pipeline to conduct social cyber security analysis. In our research, we borrow the idea of creating a pipeline of data pre-processing and analysis through the utilization of network analysis tools such as ORA and Bot Hunter, and NetMapper.



Test Set, Control Set 1, and Control Set 2

To understand if female Iranian journalists/academics/activists experienced higher harassment online, we collect datasets on female Iranian Journalists who operate outside of Iran. We also create Control Set 1 for non-Iranian journalists and Control Set 2 for male journalists of the Iranian diaspora who were also reporting on the same political events. We compare our test set against each of the two control sets to inspect how cases of harassment differed between the two. Members of the test and control sets are listed below:

Test Set	
Name	**Profession**
Negar Mortazavi	Journalists at The Independent, The Intercept, Foreign Policy | Senior Fellow CIPolicy
Azadeh Moaveni	Associate Professor at NY University | Director, Global Journalism
Tara Sepehri Far	Researcher at Human Rights Watch
Farnaz Fassihi	Journalist at New York Times
Hoda Katebi	Social media influencer covering the protest

Control Set 1	
Name	**Profession**
David Gritten	Journalist at BBS
Cora Engelbrecht	Reporter at NY Times
Kali Robinson	Reporter at Council on Foreign Relations
Sune Rasmussen	Journalist at Wallstreet Journal
Miriam Berger	Journalist Washington Post

Control Set 2	
Name	**Profession**
Ali Vaez	Adjunct Prof at Georgetown | Director of #Iran Project
Babak Dehghanpisheh	Freelance journalist, formerly at Reuters
Jason Rezaian	Writer for Washington Post
Parham Ghobadi	Reporter BBC Persia
Hadi Nili	Reporter BBC

<u>Data Preparation Data Collection:</u>

Tweets for this study were collected using the Twitter application programming interface (API) Version 2.0 and keywords related to journalists in our control and test sets. We searched for tweets that mentioned the name of each individual or their Twitter handle.

Keywords: (Negar Mortazavi OR NegarMortazavi), (Niloofar Hamedi OR NiloofarHamedi) , (Hoda Katebi OR hodakatebi), (Tara Sepehri Far OR sepehrifar), (Azadeh Moaveni OR AzadehMoaveni), (Farnaz Fassihi OR farnazfassihi), (Cora Engelbrecht OR CoraEngelbrecht), (Kali Robinson OR KaliDRobinson), (Sune Engel Rasmussen OR SuneEngel), (David Gritten OR davidgritten), (Miriam Berger OR MiriamABerger), (Ali Vaez OR AliVaez), (Babak Dehghanpisheh OR BabakDehghan), (Jason Rezaian OR jrezaian), (Parham Ghobadi OR BBCParham) and (Hadi Nili OR HadiNili).

Time period of Tweets collected: 1st Sept 2022 – 1st Oct 2022

# of Tweets: 77,226

# of Agents: 51,992

Bot Identification:

In order to identify bots in our network we use Beskow and Carley's BotHunter. BotHunter is a "random forest regression model trained on labeled Twitter data sets … developed from forensic analyses of events with extensively reported bot activity, such as the attack against the Atlantic Council Digital Forensic Research Lab in 2017".[25] Carley and Beskow explain that the Tier-1 BotHunter calculates bot probability on the following measures:[27]

i) Network-level features: Number of followers and number of friends

ii) User-level features: Screen name length, account age

iii) Tweet-level features: Timing, content

The machine learning model then produces a bot probability for each user in the network by analyzing the mentioned features.[27] In this research, we use a bot probability score of 0.7 and higher to classify a user to be a bot so that we can limit the chance of false positives in our bot categorization. This value is based on Xian Ng, Robertson's, and Carley's research, 'On the Stability of BotHunter Scores', which recommends, "a threshold value of 0.70 for Bot classification [using BotHunter]"[28]

Organization Risk Analyzer (ORA) Software:

We perform the analysis on the cleaned dataset using the network analysis software called ORA[1], a dynamic meta-network assessment and analysis tool developed by CASOS at Carnegie Mellon University.[29] The software calculates dynamic network metrics, trail metrics, and procedures in order to group nodes, identify local patterns, and compare, and contrast networks from the perspective of a dynamic meta-network.[30] In this research, we use ORA to run stance detection, identify agents who rank high on key metrics and run BEND Analysis Report.

NetMapper:

In order to extract linguistic cues from our dataset of Tweets, we use the software NetMapper. The software was presented at the 11th International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction by Carley and Reminga.[30] The linguistic cues are metrics that are required to measure key features about the author such as the agent's emotional state, general stance on a topic, etc.[23] For the purpose of this research, we will use Net Mapper to create linguistic cues so that we can feed the information into ORA software to run BEND Report – in an effort to understand what social maneuvers the agents are trying to execute in their Tweets.

[1] Available at Netanomics, https://netanomics.com/ora-commercial-version/

Identifying Tweets that Attack the Credibility of the Journalists.

After collecting our dataset of Tweets for the test set, we explored the most shared tweets mentioning each author. We found something very peculiar. We saw that the Tweets that had the most engagement for each author seemed very similar in the words used and the structure of the Tweet itself. Below is a small sample of the referenced Tweets.

> @democracynow @NegarMortazavi This is shameful and a violation of your basic ethical duties. This is who you've chosen to platform, a mouthpiece for the same morality police that killed #MahsaAmini and the same regime currently shooting protesters dead. What's your complaint procedure please? #IranLobby https://t.co/ZnMiwBarNV

> @peterson__scott @sepehrifar @hrw Dear Scott, please do your homework next time and don't promote the outrageous lies of regime mouthpieces like Sepehrifar. Reform was a sad and bloody joke we fell for 25 years ago. We want and end to this nightmare, not reform!

> 13) Former NIAC member & leading Iran apologist/lobbyist Negar Mortazavi Mortazavi is close to Zarif and, interestingly, Ilhan Omar. In September 2015 she attended an event where @HassanRouhani delivered a speech at the Hotel Hilton in New York. https://t.co/vJl2HukrC6 https://t.co/W5ZMd1ElaM

> To Negar Mortazavi, Azadeh Moaveni, Farnaz Fasihi, Hoda Katebi and other apologists and regime puppets: Iranian women are talking to you. Hear their voices and stop kissing Dictator's ass. Your days are done. We will defeat the tyrants and bring you all to justice. #MahsaAmini https://t.co/OqaMCAupvx

> @NegarMortazavi Dear world @NegarMortazavi is working for the moderate factions of the Iranian government the equally unpopular puppet group in Iran! Every word is a lie and every story has an angle. I benefit 0! day job = financier She makes money from her lies I don't #Mahsa_Amini

We see that the Tweets seemed to target the credibility of the individuals in question. These Tweets tend to call them mouthpieces, apologists, and puppets of the Iranian regime. These words seemed to be a common occurrence for each journalist. A combination of the said words seemed to be present for all individuals in the dataset. These Tweets were pushing for the sentiment that these journalists should not be taken seriously as they do not represent the interest of the Iranian population but that of the Iranian government. All in all, these Tweets seem to target the credibility of the journalist and urged others to not take them seriously.

These Tweets caught our attention because of two primary reasons. Firstly, the Tweets suggested that these journalists represent the Iranian government. However, after going through the articles published by these individuals, it was clear that most of the content they published was anti-Iranian government. None of the authors came across as pushing the regime's agenda at all. In fact, most content they published criticized the actions of the Iranian Government during the protest and extensively reported on the atrocities committed by them. Secondly, the fact that these authors were being mentioned together in Tweets raises a red flag. We initially thought this could be a result of joint publications by these authors which could have sparked a reaction against multiple authors at the same time. However, looking through their publication we found out that none of them had any joint publications in recent times. Hence co-mentioned Tweets seemed odd.

To understand this credibility-attacking sentiment we created a word list that would capture this phenomenon by identifying Tweets that used credibility-attacking words. We called this list "Credibility Attacking Wordlist" and contains words like propaganda, sabotage, apology, lobbyist, sympathizers, lie, puppet, shameful, embarrass hypocrite, whitewashed, disappointing, brainwashed, and mouthpiece. The final word list contained other derivative words which help capture the sentiment fully. The full list of words is provided in the appendix (section 1.1).

Results

Analysis 1: Identifying Cases of Credibility Attacking Tweets Contributing to Online Harassment

We used NLP methodology to identify the number of credibility-attacking Tweets for Test Set, Control Set 1, and Control Set 2 based on the word list referred to earlier. Tweets that mentioned one of the individuals from our sample and contain words from the Credibility Attacking Wordlist are considered to be credibility-attacking Tweets. In our analysis, we investigate the campaign from Sept 15th – 30th, 2022.

The aggregated results are as follows:

Table 1			
	Test Set	Control Set 1	Control Set 2
# of Credibility Attacking Tweets	7176	10	418
Credibility Attacking Tweets as % of Total Tweets	16 %	1.4 %	1.8%

Table 1 shows that the Test Set (sample of female journalists, academics, and activists) received substantially more credibility-attacking Tweets (16%) than our two control sets. These individuals experienced instances of credibility attacking harassment 11 times more than the control set that was made up of non-Iranian journalists and almost 9 times more than the sample of male Iranian journalists.

Figure 1 on the right shows how the instances of credibility-attacking Tweets aggregated for each individual in the test set. We see that these individuals were rarely Tweeted at or mentioned in credibility-attacking Tweets prior to the protests of 2022. Credibility-attacking Tweets first started on Sept 17th, 2022, for Negar Mortazavi and Hoda Katebi, the day of Amini's burial and the first day of large-scale demonstrations. Between Sept 25 – Sept 27, 2022, credibility-attacking Tweets peaked for all members of our test set.

Figure 1

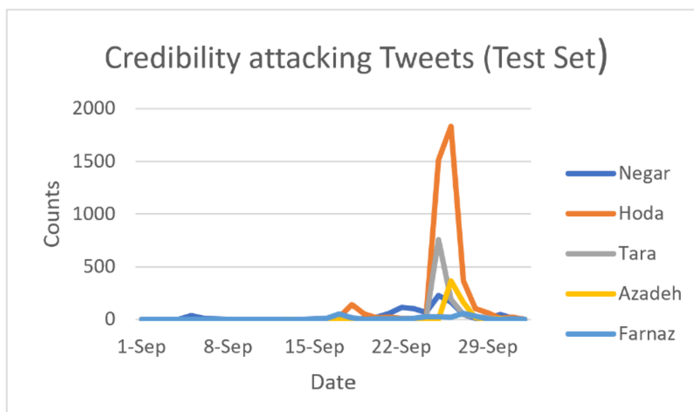


Table 2 shows that Hoda Katebi received the highest number of credibility-attacking Tweets. Between Sept 15th and Sept 30th, she experienced 4,197 instances of such Tweets. This made up 21.3 % of Tweets mentioning her in our dataset. Tara Far was the second person to be most mentioned in such Tweets and was mentioned in 1,054 instances of harassment Tweets. This made up 33.9 % of her total Tweets. The most credibility harassment Tweet as a percentage of total Tweets was experienced by Azadeh Moaveni. Almost 69% of the Tweets that mentioned Azadeh contained words that we have associated with attacking credibility.

In Control Set 1 and Control Set 2 most members hardly received any such Tweets. In Control Set 1, Cora experienced the highest number of such Tweets, but it only totaled to 5 Tweets. The highest number was experienced by Parham Ghobadi with 280 Tweets, but this only made up 2.7% of the Tweets which mentioned him.

Table2

Credibility Attacking Tweets Between Sept 15th and Sept 30th 2022

Test Set					
	Negar	Hoda	Tara	Azadeh	Farnaz
Credibility Attacking Tweets	933	4197	1054	563	292
Credibility Attacking Tweets as % of Total Tweets	4.9%	21.3%	33.9%	68.6%	11.4%
Control Set 1					
	Cora	Kali	Sune	David	Miriam
Credibility Attacking Tweets	5	0	3	0	2
Credibility Attacking Tweets as % of Total Tweets	1.4%	0%	5.3%	0%	1%
Control Set 2					
	Ali	Babak	Jason	Parham	Hadi
Credibility Attacking Tweets	105	0	16	280	17
Credibility Attacking Tweets as % of Total Tweets	3.2%	0%	0.7%	2.7%	0.3%

We performed a 2-sample t-test to find the difference in the mean number of credibility-attacking Tweets between Test Set and Control Set 1. Our findings show that the mean number of credibility-attacking Tweets for a sample of female Iranian journalists was 1407 Tweets compared to 2 Tweets for a non-Iranian journalist covering the same topic. The t-score obtained was 2.2, with 8 degrees of freedom. Comparing this t-score to the critical value for a 95% confidence level (approximately 2.306), we fail to reject the null hypothesis since the p-value is not < 0.05. However, it is important to note that the lack of significance is primarily caused by the low number of observations in our sample set. The large difference in the mean number of credibility attacking Tweets in each group suggests a noticeable difference in the harassment of Test Set and Control Set 1. The large difference in the means and the small sample size may warrant collecting more data to increase the power of the test and potentially reevaluate the results.

We then performed the Fisher's Exact Test to compare the proportion of abusive Tweets experienced by journalists in the test and control sets, as an alternative to the t-test. This decision was made due to the small sample sizes in both sets, which can lead to unreliable results when using a t-test. The results of the Fisher's Exact Test revealed a statistically significant difference in the proportion of credibility-attacking Tweets experienced by journalists in the test and control sets. With a p-value of less than 2.2e-16, which is practically 0, we have strong evidence to reject the null hypothesis and accept the alternative hypothesis that the true odds ratio is not equal to 1. The estimated odds ratio of 12.33 indicates that the odds of encountering credibility-attacking Tweets for journalists in the Test Set are approximately 12.34 times higher than the odds for journalists in the control set 1. Furthermore, the 95% confidence interval for the odds ratio ranges from 6.663399 to 25.857981, providing additional support for the conclusion that there is a significant difference in the likelihood of encountering credibility-attacking Tweets between the two groups of journalists. This finding highlights the disparity in the online experiences of the journalists studied.

We also performed a 2-sample t-test to find the difference in the mean of credibility-attacking Tweets between Test Set and Control Set 2. Our findings show that the mean number of credibility-attacking Tweets for the sample of female Iranian journalists was 1407 Tweets compared to 83.6 for male Iranian journalists covering the same topic. We see a similar lack of significance in our results, but this is due to the small sample size again. The stark difference between the mean number of credibility-attacking Tweets experienced by the two groups warrants attention.

Next, we run an OLS regression to see if there is an association between being female and Iranian to the amount of credibility attacking Tweets experienced between 15th Sept and 30th Sept 2022.

# of Credibility attacking Tweets = B0 + B1Female + B2Iranian + E

Our results show that being female is associated with experiencing 895 more Tweets of credibility attacking compared to being male, holding all factors constant. Our results also show that being Iranian is associated with experiencing 833 more Tweets of credibility attacking compared to being non-Iranian, holding all factors constant. Here we experience low p-value for dependent variables. This again is a result of a low number of observations in each group. However, significant differences in averages of each group and large coefficients for the regression model make a case for possible large differences in harassment experienced as a result of being female and Iranian. Increasing the number of individuals in each sample could potentially result in increased statistical significance for analysis without significant changes in the coefficients.

<u>Presence of Bot Accounts in Credibility Attacking Tweets:</u>

We used Beskow and Carley's Tier 1 bot hunter algorithm to identify the bot probability of each account in our dataset. Below is a trend of bot participation in Test Set, Control Set 1 and Control Set 2. We classify each account as bot if it has a bot probability higher than 0.7.

Figure 2 shows how the total number of bots in each network changes with time. We see very minimal bot presence for control set 1 and 2. For our Test Set, the total number of bots in credibility attacking Tweets peaks on 25th Sept 2022.

Figure 2

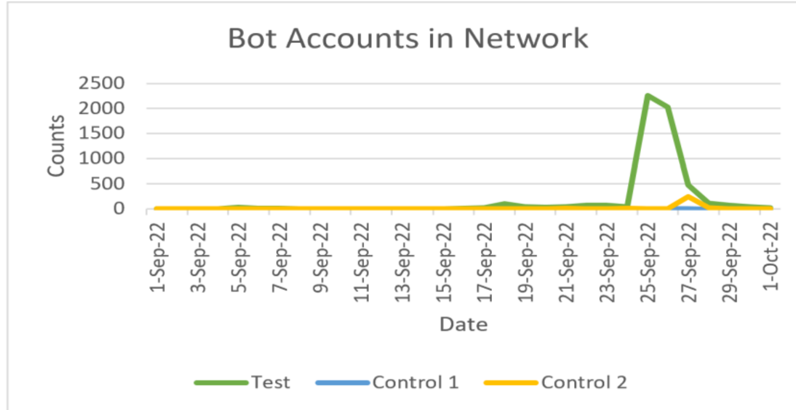


Table 3 shows that in the network of credibility-attacking Tweets experienced by the Test Set (between Sept 15 – 30, 2022), there were 3,796 bots identified. This suggested that 79.3% of all accounts in the credibility-attacking network were bot accounts. These 3,796 bots were responsible for 78.3% of the Tweets which were attacking the credibility of female journalists from the Iranian diaspora. By comparison, when we look at the sample of male journalists of the Iranian diaspora. We see that there were 344 bots present in the network of credibility-attacking Tweets which made up 86% of all accounts. These bots contributed to creating 77.4% of credibility-attacking Tweets. The sample of non-Iranian journalists experienced harassment by 2 bot accounts only which made up 16% of the accounts in the network.

Table 3

Bot Accounts in the Network (Sep 15 -30, 2022)			
	Test Set	Control Set 1	Control Set 2
# of bot accounts in the network	3796	2	344
% of accounts that are bots	79.3%	16.6%	86%
Tweets by Bot Accounts in the Network			
# of Tweets by bot accounts	5553	2	358
% of accounts that are bots	78.3%	16.6%	77.4%

BEND Analysis of credibility attacking Tweets:

Through BEND Framework analysis we were able to divide these credibility-attacking Tweets by the bot accounts into 16 different categories of narrative and network maneuvers. These maneuvers provide a granular understanding of how the accounts were trying to manipulate the conversation that was taking place in the network and the structure of the network itself.

Our analysis showed that most Tweets were categorized in the Dismiss category. About 80.4% (4,500 Tweets) of the bot-generated credibility-attacking Tweets were classified as Dismiss Tweets. Dismiss maneuver is one of the categories of the negative narrative maneuver under the BEND framework. Dismiss Tweets seek to reject the narrative that is being propagated by the target Tweeter. In our dataset, we saw that most of the Tweets belonging to this category were Tweets that were a reply to the original Tweet posted by the author. These dismissed Tweets referred to the earlier Tweets by the author and encouraged the readers to not take the journalist's Tweets seriously since they are mouthpieces, apologists, and lobbyists for the Iranian government. The BEND framework's identification of Dismiss as the most prevalent maneuver category is significant in the context of credibility-attacking Tweets, as it shares many characteristics with such Tweets. This is a positive development and reinforces the framework's usefulness in detecting such Tweets.

The second most common maneuver by bots in Test Set was Dismay and 66.27% of the Tweets were categorized under this maneuver. Dismay is categorized as a narrative maneuver that seeks to elicit negative emotions such as sadness and anger toward the target. The narrative focused on harassing the journalists for being an agent of the government and pointed criticism towards them for their insincere work. These Tweets were an intrinsic part of the bot campaign to convince readers to also dislike the individuals in our Test Set. The third most common maneuver adopted by the bot accounts was Nuke. Nuke is a form of network maneuver that seems to change the structure of a topic-focused community. In our sample Test Set, 65.5% of the Tweets were categorized under this category. Tweets under this category were directed towards the journalists in our Test Set and urged the readers to not consider them as a representative of the movement. Nuke Tweets also frequently mentioned the employers of the journalists such as the NY Times, CNBC, Intercept, Human Rights Watch, and Georgetown University, and suggested that they should be ashamed for employing these journalists. These Nuke Tweets were hoping to not only attack the credibility of the authors but also instruct readers to disassociate themselves from these journalists – in an effort to reduce the following of these journalists in specific topic-focused communities.

We also compare the BEND maneuvers of harassment Tweets by bot accounts vs that by organic users for our test set. The results of distribution by BEND category for bot and organic accounts are available in the appendix (section 1.6). Our analysis showed that harassment by bot accounts primarily focused on negative maneuvers. The negative maneuvers were primarily focused on limiting the discussion the female Iranian journalists were having and reducing their reach in the network. Credibility harassment by organic accounts used mostly positive maneuvers that were trying to only show their displeasure with the authors but not necessarily tried to limit the conversation or change the structure of the network in an adverse way.

Analysis 2: Identifying Cases of Abusive Tweets Contributing to Online Harassment

After having explored cases of credibility-attacking Tweets against the female journalists of the Iranian diaspora, we also tried to track abuse that was experienced by journalists and activists in our samples. Abusiveness can be described as Tweets that contain offensive words when referring to the Tweeters. Abusive Tweets contain offensive language, which is insulting, and use curse words or words associated with rude and vulgar language. These abusive Tweets are identified using NetMapper to find instances of abusive words in Tweets that mentioned journalists in our samples.

We used NLP to identify the number of abusive Tweets for Test Set, Control Set 1, and Control Set 2 using linguistic cues created on our samples.

The results of the trends are as below:

Table 4			
	Test Set	Control Set 1	Control Set 2
# Abusive Tweets	9099	122	1055
Abusive Tweets as % of Total Tweets	19.93%	17.96%	4.74%

Table 4 shows that the group of Iranian female journalists, academics, and activists in our study were targeted with significantly more abusive Tweets (19.93%) compared to the control set of male journalists from the Iranian diaspora (4.74%). Specifically, our test set faced almost 5 times more instances of abusive Tweets. We see that the amount of abuse received by non-Iranians totaled 17.96% which was quite similar to that experienced by Test Set 1 however there were only 122 Tweets.

We look at the data more granularly to identify who were the individuals who received the most abuse in our dataset between Sept 15th -20th, 2022. In our Test Set, we see that most abuse was received by Negar Mortazavi who was mentioned in 6,526 abusive Tweets (34% of the Tweets she was mentioned in), followed by Hoda Katebi who was mentioned in 1,772 abusive Tweets (8.98% of the Tweets she was mentioned in). When looking at our Control Set 1 we found out that 104 abusive Tweets out of 137 were received by Cora Engelbrecht. Cora Engelbrecht is a reporter at NY Times, and she regularly co-publishes articles with Farnaz Fasih who is one of the female journalists of the Iranian diaspora and part of our Test Set as well. Cora started to experience abusive Tweets as she was closely associated with Farnaz as well.

Table 5					
Test Set					
	Negar	Hoda	Tara	Azadeh	Farnaz
# of abusive Tweets	6,526	1772	433	7	237
Abusive Tweets as % of Total Tweets	34.35%	8.98%	13.9%	0.85%	10.69%
Control Set 1					
	Cora	Kali	Sune	David	Miriam
# of abusive Tweets	104	15	6	3	9
Abusive Tweets as % of Total Tweets	28.49%	0	10.7 %	9.37%	4.26%
Control Set 2					
	Ali	Babak	Jason	Parham	Hadi
# of abusive Tweets	258	6	83	443	265
Abusive Tweets as % of Total Tweets	7.86%	6.52%	3.69%	4.25%	4.27%

We conducted a 2-sample t-test to examine the difference in the mean number of abusive Tweets between Test Set and Control Set 1. Our analysis revealed that the average number of abusive Tweets for a group of female Iranian journalists was 1,802 Tweets, compared to 24 Tweets for a non-Iranian journalist reporting on the same subject. The calculated t-value was 0.3930, which, when compared to the critical value of approximately 2.306 at a 95% confidence level. Again, as a result of the small sample size the results were statistically insignificant. Nonetheless, the substantial disparity in the means and the limited sample size suggest that gathering more data to enhance the test's power could be beneficial, potentially leading to a reevaluation of the results.

Similarly, we carried out a 2-sample t-test to explore the difference in the mean number of abusive Tweets between the Test Set and Control Set 2. Our findings indicated that the average number of abusive Tweets for a sample of female Iranian journalists was 1,802 Tweets, as opposed to 211 Tweets for non-Iranian journalists covering the same topic. The t-value calculated was 1.0473 with 8 degrees of freedom, which, when compared to the critical value of approximately 2.306 at a 95% confidence level, was insufficient to reject the null hypothesis. However, the considerable difference in the means and the small sample size imply that collecting additional data to improve the test's power may be advantageous, potentially prompting a reassessment of the outcomes.

We further break down the abusive Tweets for each group into more granular sub-categories to get a better understanding of what was the abuse affiliated with. Table 6 shows the number of abusive Tweets that belonged to each sub-category for each test and control group. Our results show that Test Set experienced abusive content in all subcategories that was substantially more than that experienced by the two control groups. The highest abuse received by female journalists of the Iranian diaspora was in the categories of gender-based abuse, Job-based abuse, and violence-propagated abuse. Gender and job being the top category of abuse further asserts our claim that the individuals in our Test Set face abuse for being female journalists. We also note that the average negative score for abusive Tweets was more negative for our Test Set (-0.71) compared to the Control Set 1 (-0.59) and Control Set 2 (-0.63). This suggests that on average each abusive Tweet that was experienced by Test Set had more abusive words per Tweet than abusive Tweets experienced by other groups.

Table 6

Sub-categories of Abusive Tweets

	Test Set	Control Set 1	Control Set 2
Political	88	5	48
gender	7243	34	471
Religion	78	0	26
Race/nationality	1658	20	331
Job	7591	51	457
Violence	6746	133	482
Avg Positive	0.33	0.45	0.3
Avg Negative	-0.71	-0.59	-0.63

Presence of Bot Accounts in Credibility Attacking Tweets:

Our findings indicate that in our network of abusive Tweets, Test Set contained 7,336 bot accounts, accounting for 89.8% of all accounts in the abusive network. These 7,336 bots contributed to 88.7% of the Tweets harassing female journalists from the Iranian diaspora with abusive content. In contrast, in the sample of male Iranian diaspora journalists, there were 830 bots involved in the abusive harassment campaign, making up 68.8% of all accounts. These bots were responsible for 77.4% of the abusive Tweets. The sample of non-Iranian journalists experienced harassment from only 86 bot accounts, which constituted 69% of the accounts in the network. The overall trend of bot participation shows that the bot participation started on Aril 14th, a day after the death of Masha Amini and continued to remail in the Test Set for the extent of the period we are analyzing.

Figure 3

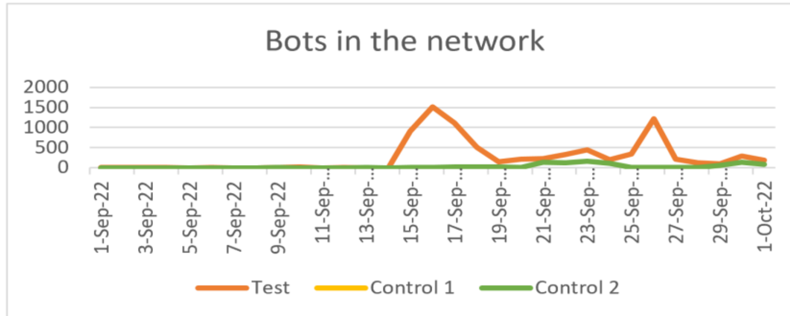


Table 7			
Bot Accounts in the Network			
	Test Set	Control Set 1	Control Set 2
# of bot accounts in the network	7336	86	830
% of accounts that are bots	89.79%	69.35%	68.82%
Tweets by Bot Accounts in the Network			
# of Tweets by bot accounts	8078	97	877
% of accounts that are bots	88.7%	69.2%	77.4%

Discussion:

The analysis of harassment campaigns against female journalists of the Iranian diaspora presented in this paper highlights the presence of targeted credibility-attacking and abusive Tweets against these individuals. Our findings demonstrate a significantly higher proportion of credibility-attacking and abusive Tweets against female journalists of Iranian origin compared to the control sets of male journalists of the Iranian diaspora and non-Iranian journalists between September 15th – 30th, 2022. This suggests that female journalists from the Iranian diaspora may be targeted specifically because of their gender and professional affiliation.

Firstly, in our analysis, we divided harassment into two distinct categories. One category of harassment was credibility-attacking Tweets while the other category of harassment was Tweets with abusive language. In our analysis, we saw that female journalists/activists experienced significantly higher levels of credibility attacking Tweets compared to the sample of non-Iranian journalists and male Iranian journalists. 16% of Tweets received by female Iranian journalists were such Tweets. For most journalists in our Test Set, levels of harassment were higher than this. For e.g., 68.6% of Tweets mentioning Azadeh Moaveni, 33.9% mentioning Tara Far, and 21.3% mentioning Hoda Katebi were credibility-attacking Tweets. The total percentage of 16% is low primarily due to the fact that we have Hoda Katebi and Negar Mortazavi in our data sets who experience a lot of Tweets directed at them which are not-necessarily credibility-attacking Tweets, thus bringing the overall percentage of such Tweets down in our dataset. In comparison non-Iranian journalists and male journalists of the Iranian diaspora only experienced 1.4% and 1.8% of credibility attacking Tweets respectively. Running a regression suggested that being female is associated with 859 more cases of credibility-attacking Tweets and being Iranian is associated with 833 more such Tweets. Additionally, a vast proportion of the credibility-attacking harassment was orchestrated by Bot accounts. In the network of Iranian females, we saw 79.3% of accounts that were propagating a narrative of challenging journalists' credibility were Bot accounts. These Bot accounts primarily took part in negative narrative and network maneuvers. The top maneuvers in this category of harassment consisted of Dismiss maneuver (80.4 % of Tweets), the Dismay Maneuver (66.3% of Tweets), and the Nuke maneuver (65.5% of Tweets). Within these maneuvers, the primary objective of bot accounts was to reject the narrative that is being propagated by the female Iranian journalists, to elicit negative emotions such as anger toward the target journalists. These Tweets further urged the audience and employers of the journalist to disassociate themselves from these individuals.

Secondly, when we look at abusive Tweets, we found that the sample of female Iranian journalists experienced a significantly higher number of such Tweets as well. Between September 15th – 30th, 2022, we saw that female journalists experienced 6,526 abusive Tweets which made up 19.9% of the Tweets in which these authors were mentioned. In comparison, male Iranian journalists experienced 1,055 such Tweets which made up 4.7% of Tweets in the network. Non-Iranian journalists had 17.9% of abusive Tweets in their network which seemed like a high proportion but only totaled 122 Tweets. This is because non-Iranian journalists did not receive a large number of Tweets (abusive and non-abusive) in the period. Interestingly Cora Engelbrecht received the highest number of abusive Tweets in the sample of non-Iranian Journalists. Close to 30% of the Tweets which mentioned her were abusive Tweets. This high percentage for Cora could be attributed to her working with Farnaz Fassihi, a female journalist of the Iranian diaspora who is one of the members of our Test Set too. Since Cora and Farnaz are co-authors on a number of articles related to Iran published by the New York Times, it was not surprising to see that Cora also received high abuse.

Additionally, we also saw that female Iranian journalists received high abuse in all subcategories of abuse. The highest abuse for these individuals was in the categories of Gender and Profession based abuse. This further reiterates our concern that the group of journalists that we were interested in tracking abuse for were primarily targeted for their gender as a female and working in the journalism profession. The abusive Tweets for the female Iranians were also more negative as each Tweet had a higher average negative score than the abusive Tweets experienced by the other two groups. Our analysis also showed that the abusive campaign for these journalists was propagated by Bot accounts. We saw that close to 90% of the accounts that were taking part in abusing female Iranian journalists were Bot accounts.

Limitations and Recommendations:

1) Small Sample Size: One of the main limitations of this study is the small sample size. Each of the three groups in our study consisted of only five journalists/activists, which may not adequately represent the experiences of the broader population of female journalists of the Iranian diaspora, male Iranian diaspora journalists, and non-Iranian journalists. The small sample size may have affected the statistical power of our analyses, leading to the inability to reject null hypotheses in some cases, despite observing considerable differences in the means between groups.

   We recommend for future studies look at larger samples of journalists. This could potentially yield more conclusive results and provide a better understanding of the experiences of each group. Additionally, expanding the sample size could help to identify more diverse experiences within each group, as well as detect patterns and trends that may not be apparent in smaller samples. Therefore, future research should consider utilizing larger sample sizes to enhance the robustness of the findings and better represent the experiences of the target population.

2) Broder study on women journalists more than just Iran: One notable limitation is the narrow focus on female Iranian journalists, which may not fully represent the experiences of women journalists from other nationalities. Expanding the sample set to include female journalists from various nationalities would provide a more comprehensive understanding of online harassment patterns and the extent of the problem. This broader sample would enable researchers to identify any commonalities or differences in harassment experienced by women journalists across the globe. Additionally, it could help in understanding the role of cultural, political, and social contexts in shaping the nature and intensity of online harassment. We recommend future research consider a more diverse sample set to gain a better understanding of the global phenomenon of online harassment against female journalists. This expanded analysis would be instrumental in designing effective interventions and strategies to address and mitigate the issue of online harassment and create a safer online environment for women journalists worldwide.
3) Controlling for NIAC membership: The National Iranian American Council is an NGO based in Washington D.C. that advocates for American Iranian issues. In the past NIAC has been known to receive harassment by some claiming that it does not represent the interest of all Iranians in the diaspora. In our study, we had members of both NIAC and those who were not. Not controlling for membership to this organization will result in not realizing if this factor was having a significant

impact on levels of harassment experienced by those in samples of female Iranian journalists and male Iranian journalists. Ideally, we would have accounted for this factor, but it seems that membership to NIAC is difficult to identify. Sometimes individuals who are not official members are also known to be close to the organization and receive online harassment. Our recommendation would be to have an expert who understands the nuances of this organization and understand the grievances of different sub-groups towards different members of this organization. This will help in the classification of who is a known associate of NIAC so that this factor can be controlled for in future works.

Appendix

1.1: Word List of Credibility attacking Tweets

Category: Credibility attacking words

Key Words: Propaganda, sabotage, apology, lobbyist, sympathizers, lie, puppet, shameful, embarrass, hypocrite, whitewashed, disappointing, brainwashed, and mouthpiece.

Word List:

propaganda
Propoganda
propagandist
propagandists
propagandize
propagandizing
misinformation
disinformation
fakenews
sabotage
Sabotaged
Sabotages
sabotaging
undermine
undermined
underminds
undermining
apology
apology
apologist
apologists
apologies
apologizing
apologized
apologizes
lobbyist
lobbying
lobby
lobbies
IranLobby
lobbists
lobbied
sympathizers
sympathizer
Sympathized
sympathize
sympathizes
sympathizing
pro-regime
lie
liar
lying
lied
liars
mislead
misleading
misleads
misleaders
puppet
puppets
puppeteer
pawn
pawns
shameful
ashamed
shameless
shame
shamed
embarrassment
embarrassing
ebbarassed
embarrasses
hypocrite
hypocrites
hypocritical
hypocrisy
whitewashed
whitewashing
whitewashes
whitewashed
disappointing
dissapoints
dissapointed
dissapoint
brainwashed
barianwashing
brainwashed
brainwashes
brainwash
mouthpiece
mouthpieces

1.2: Difference in mean number of credibility attacking Tweets for Test Set and Control Set 1

```
        Welch Two Sample t-test

data:  set1 and set2
t = 1.9793, df = 4, p-value = 0.1189
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -566.1536 3377.7536
sample estimates:
mean of x mean of y
   1407.8       2.0
```

1.3: Difference in mean number of credibility attacking Tweets for Test Set and Control Set 2

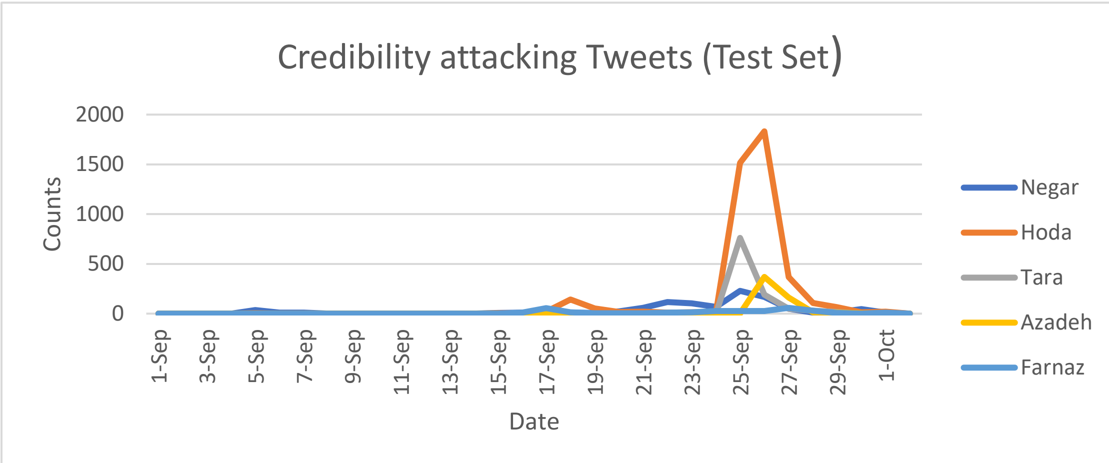
```
        Welch Two Sample t-test

data:  set1 and set2
t = 1.8594, df = 4.0436, p-value = 0.1357
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -644.739 3293.139
sample estimates:
mean of x mean of y
   1407.8      83.6
```

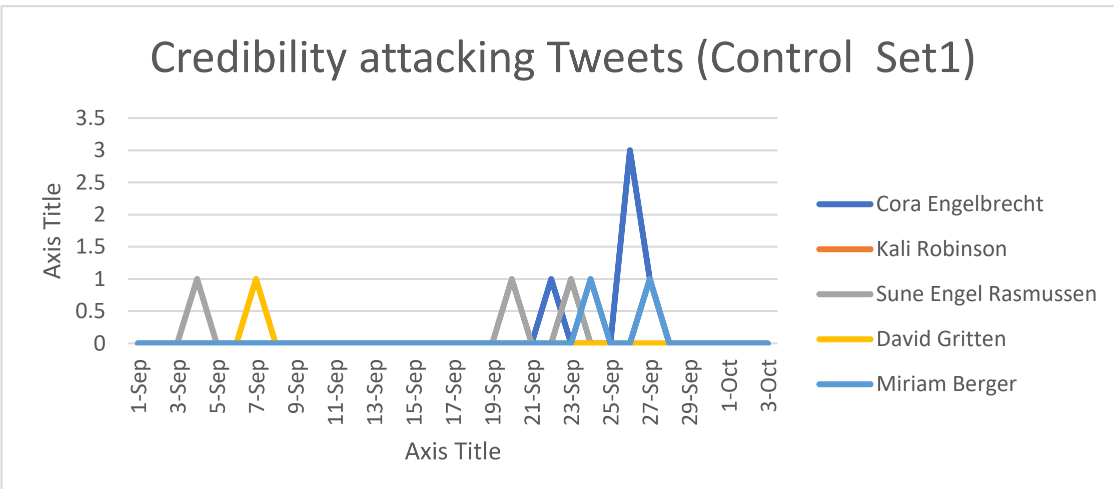
1.4: Regression with counts of Credibility harassment as dependent variable

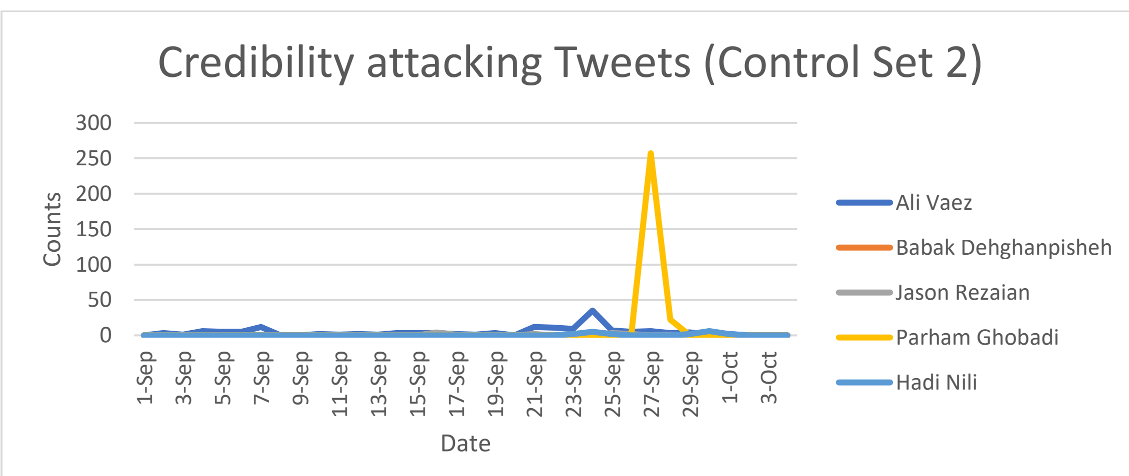
cred_harrasment = B0 + B1Female + B2Iranian + E

```
===============================================
                        Dependent variable:
                    ---------------------------
                         cred_harrasment
-----------------------------------------------
Female                       895.000
                            (510.343)

Iranian                      833.200
                            (540.096)

Constant                    -535.000
                            (535.252)

-----------------------------------------------
Observations                   15
R2                           0.294
Adjusted R2                  0.176
Residual Std. Error      981.663 (df = 12)
F Statistic             2.494 (df = 2; 12)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

1.5: Trends of Credibility Attacking Tweets

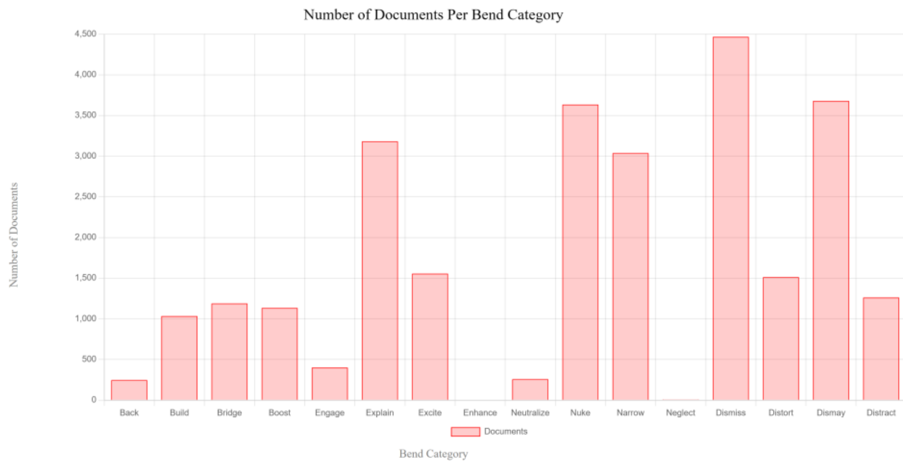




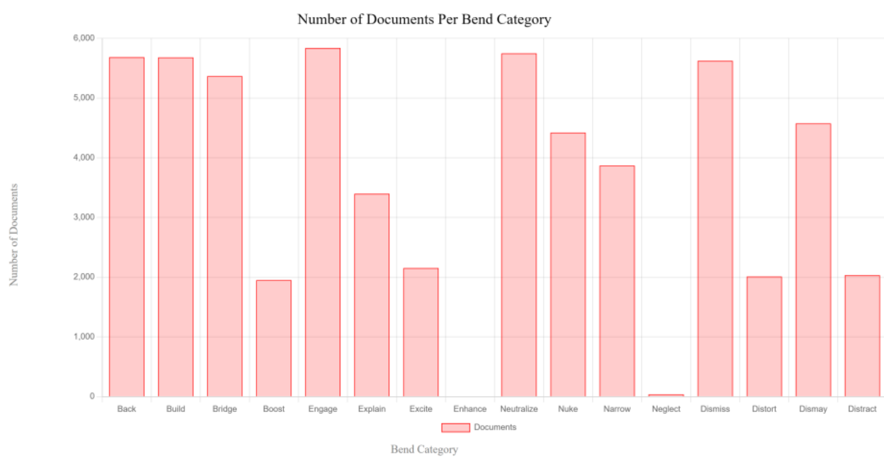


1.6: BEND categorization for credibility attacking Tweets by bots and organic account

By Bot Accounts



By Organic Accounts



References:

[1] “Iran: Deadly Crackdown on Protests against Mahsa Amini's Death in Custody Needs Urgent Global Action.” Amnesty International, September 23, 2022. https://www.amnesty.org/en/latest/news/2022/09/iran-deadly-crackdown-on-protests-against-mahsa-aminis-death-in-custody-needs-urgent-global-action/.

[2] “Iran: Mahsa Amini's Father Accuses Authorities of a Cover-Up.” BBC News. BBC, September 22, 2022. https://www.bbc.com/news/world-middle-east-62998231.

[3] “Iran: Death Sentences against Protesters.” Human Rights Watch, December 13, 2022. https://www.hrw.org/news/2022/12/13/iran-death-sentences-against-protesters.

[4] “In Iran, Crackdowns on Dissent and Journalism.” Columbia Journalism Review. Accessed March 18, 2023. https://www.cjr.org/the_media_today/iran-protests-niloufar-hamedi-mahsa-amini.php.

[5] “Witch Hunt in Iran: Grave Concerns for Journalists Niloofar Hamedi and Elahe Mohammadi.” Center for Human Rights in Iran, November 18, 2022. https://iranhumanrights.org/2022/11/witch-hunt-in-iran-grave-concerns-for-journalists-niloufar-hamedi-and-elahe-mohammadi/.

[6] Staff. “Iranian Photojournalist Yalda Moaiery Arrested during Anti-State Protests .” Committee to Protect Journalists, September 20, 2022. https://cpj.org/2022/09/iranian-photojournalist-yalda-moaiery-arrested-during-anti-state-protests/.

[7] Khan, Sahar. “Women Journalists Face Increasing Danger in Iran.” Inkstick. Inkstick Media, September 30, 2022. https://inkstickmedia.com/women-journalists-face-increasing-danger-in-iran/.

[8] Ghobadi, Parham. “Why Reporting on Iran Comes at a Heavy Price.” BBC News. BBC, January 12, 2023. https://www.bbc.com/news/world-middle-east-64222261.

[9] “Online Harassment against Women Journalists in the Iranian Diaspora.” ARTICLE 19, October 19, 2021. https://www.article19.org/resources/online-harassment-against-women-journalists-in-the-iranian-diaspora/.

[10] Ghobadi, Parham. “Why Reporting on Iran Comes at a Heavy Price.” BBC News. BBC, January 12, 2023. https://www.bbc.com/news/world-middle-east-64222261.

[11] Keskin, Enes. “United States-Iran: CFIWJ Stands in Solidarity with Farnaz Fasihi Who Was Targeted through Vicious Online Trolling.” Coalition For Women in Journalism. Coalition For Women in Journalism, December 8, 2021. https://www.womeninjournalism.org/threats-all/us-iran-cfwij-stands-in-solidarity-with-farnaz-fasihi-who-was-targeted-through-vicious-online-trolling.

[12] Kundu, Priyanka, and Md Mahbubul Haque Bhuiyan. "Online harassment of female journalists in Bangladesh: Forms, reactions, and consequences." In Handbook of research on discrimination, gender disparity, and safety risks in journalism, pp. 143-166. IGI Global, 2021.

[13] Wotanis L and McMillan L (2014) Performing gender on YouTube. Feminist Media Studies 14(6): 912–928.

[14] Cole KK (2015) ‘It’s like she’s eager to be verbally abused’: Twitter, trolls, and (en)gendering disciplinary rhetoric. Feminist Media Studies 15(2): 356–358.

[15] Chen, Gina Masullo, and Paromita Pain. "Normalizing online comments." Journalism Practice 11, no. 7 (2017): 876-892.

[16] Kundu, Priyanka, and Md Mahbubul Haque Bhuiyan. "Online harassment of female journalists in Bangladesh: Forms, reactions, and consequences." In Handbook of research on discrimination, gender disparity, and safety risks in journalism, pp. 143-166. IGI Global, 2021.

[17] Demos (2014). Demos: Male celebrities receive more abuse on Twitter than women. Retrieved from https://demos.co.uk/press-release/demos-malecelebrities-receive-more-abuse-on-twitter-thanwomen-2/

[18] Martin, Fiona R., and Colleen Murrell. "You need a thick skin in this game: Journalists' attitudes to resilience training as a strategy for combatting online violence." australian journalism review 42, no. 1 (2020): 93-111.

[19] Chen, Gina Masullo, Paromita Pain, Victoria Y. Chen, Madlin Mekelburg, Nina Springer, and Franziska Troger. "'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists." Journalism 21, no. 7 (2020): 877-895.

[20] Online Harassment of Journalists: Attacks of the trolls. Reporters Without Borders. 21 March 2023. https://rsf.org/sites/default/files/rsf_report_on_online_harassment.pdf

[21] Rego, Richard. "Changing forms and platforms of misogyny: Sexual harassment of women journalists on twitter." Media watch 9, no. 3 (2018): 472-485.

[22] Becky Gardiner (2018) "It's a terrible way to go to work:" what 70 million readers' comments on the Guardian revealed about hostility to women and minorities online, Feminist Media Studies, 18:4, 592-608, DOI: 10.1080/14680777.2018.1447334

[23]Carley K. M. (2020). Social cybersecurity: an emerging science. Computational and mathematical organization theory, 26(4), 365–381. https://doi.org/10.1007/s10588-020-09322-9

[24] National Academies of Sciences Engineering and Medicine (U.S.). (2019). A decadal survey of the social and behavioral sciences : a research agenda for advancing intelligence analysis. National Academies Press. Retrieved September 7 2022 from (chaper 6) https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5811670

[25] Blane, J. T., Bellutta, D., & Carley, K. M. (2022). Social-cyber maneuvers during the COVID-19 vaccine initial rollout: Content analysis of tweets. Journal of Medical Internet Research, 24(3). https://doi.org/10.2196/34040

[26] Uyheng, J., Magelinski, T., Villa-Cox, R., Sowa, C., & Carley, K. M. (2019). Interoperable pipelines for social cyber-security: Assessing twitter information operations during NATO Trident Juncture 2018. Computational and Mathematical Organization Theory, 26(4), 465–483. https://doi.org/10.1007/s10588-019-09298-1

[27] Beskow, D.M., & Carley, K.M. (2018). Bot-hunter: A Tiered Approach to Detecting & Characterizing Automated Activity on Twitter.

[28] Ng LHX, Robertson DC, Carley K. On the stability of BotHunter scores. 2021 Jul 12 Presented at: Center for Informed Democracy & Social cybersecurity (IDeaS) Conference; July 12-13, 2021; Virtual event URL: https://www.cmu.edu/ideas-socialcybersecurity/events/conference-archive/archive-conference-2021/conference-papers2021/conference-2021-paper-17-bothunter.pd

[29] Kathleen M. Carley (2017), "ORA: A Toolkit for Dynamic Network Analysis and Visualization." In Reda Alhajj and Jon Rokne (Eds.) Encyclopedia of Social Network Analysis and Mining, Springer

[30] Carley LR, Reminga J, Carley KM. ORA & NetMapper. 2018 Presented at: 11th International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation; July 10-13, 2018; Washington, DC.