

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

---

12-2020

### Digital social listening on conversations about sexual harassment

Xuesi SIM

Ern Rae CHANG

Yu Xiang ONG

Jie Ying YEO

Christine Bai Shuang YAN

*See next page for additional authors*

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)



Part of the [Asian Studies Commons](#), [Databases and Information Systems Commons](#), and the [Social Media Commons](#)

---

#### Citation

SIM, Xuesi; CHANG, Ern Rae; ONG, Yu Xiang; YEO, Jie Ying; YAN, Christine Bai Shuang; CHOY, Eugene Wen Jia; and SHIM, Kyong Jin. Digital social listening on conversations about sexual harassment. (2020). *2020 IEEE International Conference on Big Data: Virtual conference, December 10-13: Proceedings*. Research Collection School Of Computing and Information Systems.

Available at: [https://ink.library.smu.edu.sg/sis\\_research/5650](https://ink.library.smu.edu.sg/sis_research/5650)

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylds@smu.edu.sg](mailto:cherylds@smu.edu.sg).

---

**Author**

Xuesi SIM, Ern Rae CHANG, Yu Xiang ONG, Jie Ying YEO, Christine Bai Shuang YAN, Eugene Wen Jia CHOY, and Kyong Jin SHIM

# Digital Social Listening on Conversations About Sexual Harassment

Xuesi Sim  
School of Information Systems  
Singapore Management  
University  
Singapore  
xuesi.sim.2017@smu.edu.sg

Ern Rae Chang  
School of Information Systems  
Singapore Management  
University  
Singapore  
erchang.2017@smu.edu.sg

Yu Xiang Ong  
School of Information Systems  
Singapore Management  
University  
Singapore  
yuxiang.ong.2016@smu.edu.sg

Jie Ying Yeo  
School of Information Systems  
Singapore Management  
University  
Singapore  
jieying.yeo.2017@smu.edu.sg

Christine Bai Shuang Yan  
School of Information Systems  
Singapore Management University  
Singapore  
ybayan.2017@smu.edu.sg

Eugene Wen Jia Choy  
School of Information Systems  
Singapore Management University  
Singapore  
eugene.choy.2016@smu.edu.sg

Kyong Jin Shim  
School of Information Systems  
Singapore Management University  
Singapore  
kjshim@smu.edu.sg

**Abstract**—In light of the #MeToo movement and publicized sexual harassment incidents in Singapore in recent years, we built an analytics pipeline for performing digital social listening on conversations about sexual harassment for AWARE (Association of Women for Action and Research). Our social network analysis results identified key influencers that AWARE can engage for sexual harassment awareness campaigns. Further, our analysis results suggest new hashtags that AWARE can use to run social media campaigns and achieve greater reach.

**Keywords**— sexual harassment, analytics, social network analysis, social media, #MeToo

## I. INTRODUCTION

AWARE (Association of Women for Action and Research) Singapore is a non-profit organization established in 1985 that advocates gender equality. In response to sexual violence cases in Singapore, AWARE has launched the Sexual Assault Care Centre (SACC) which is Singapore’s first and only known specialized care center for survivors of sexual violence. In recent years, they have initiated multiple physical campaigns such as “We Can!” campaign, a two-year Aim for Zero campaign and online global campaigns such as #16DaysSG. Sexual harassment has been a worrying issue in Singapore. According to a recent survey, nearly half of the students and recent graduates from the two largest universities in Singapore experienced sexual misconduct on campus, yet most cases go unreported [1]. Last year, a National University of Singapore (NUS) student shared her own ordeal online, and it was highly publicized in social media [2]. Her incident went viral and sparked discussions on how we could prevent and better respond to sexual harassment in Singapore.

In response to this problem, another local university – Singapore Management University (SMU) formed a first-

response unit called Voices@SMU which provides support services to victims and perpetrators. The university created a “Supporting a Harassment-Free Environment” module in their Learning Management System and planned to organize campus campaigns to raise greater awareness on this issue. Over the last few years, AWARE has also been actively conducting physical campaigns and online global campaigns such as #16DaysSG to educate the public on the issue of sexual violence.

Building onto AWARE’s existing initiatives, we aim to help them plan and execute their public campaigns through a streamlined social analytics pipeline. Our problem statement is “How might we expand the online reach of our client’s existing awareness efforts on sexual harassment?” In this project, we define sexual harassment as “unwelcome requests for sexual favors and verbal, non-verbal or physical misconduct of a sexual nature” [3]. In determining whether a behavior is considered sexual harassment, we decided to base that on the perception of the receiver and not the intention of the sender [3]. We aim to address the following questions:

1. How can AWARE reach out to wider audience? (AWARE aims to achieve 1000 pledges for AimForZero Campaign by Nov 2020)
2. How to drive greater awareness on this topic?
3. Which social media platforms have most discussion on sexual harassment?

## II. DATA COLLECTION

To expand the reach and awareness efforts on sexual harassment online, data was collected from Instagram and Twitter.

### A. Instagram

For Instagram social media data, a total of 80,000 posts were scraped, and the data duration spanned over three years (2017 - 2019). The public Instagram API was used to scrape textual data which were subsequently stored in respective hashtag's .json files.

### B. Twitter

We identified sexual harassment related tweets using related search terms (Fig. 1). As the twitter API could only crawl up to about seven past days of data, we chose to use Phantom Buster to web scrape all the relevant tweets that had been posted over the years. A total of 1,262 tweets and 12 features were scraped that spanned over a period of 11 years (2009 - 2020). We have also leveraged the Twitter API and Tweepy package in Python to obtain each user's follower count.

```
#metoo, #hetoo, #voyeurism, #sexualharassment,
#sexualmisconduct, #monicabaey, #aimforZeroSG,
#sexualassault, #sexualviolence, #inappropriatetouching,
#maleharassment, #rape, #inappropriate comment
```

Fig. 1. Twitter - Search Terms

## III. SOCIAL NETWORK ANALYSIS

### A. Instagram

In many social media platforms, hashtagging is a powerful feature for promoting and spreading ideas. The many short hashtags reveal user's sentiments and preferences. In our analysis, we investigated co-occurring hashtags in Instagram posts. Scoping the data down to only the targeted hashtags (related to sexual harassment), each node represents a hashtag in the network.

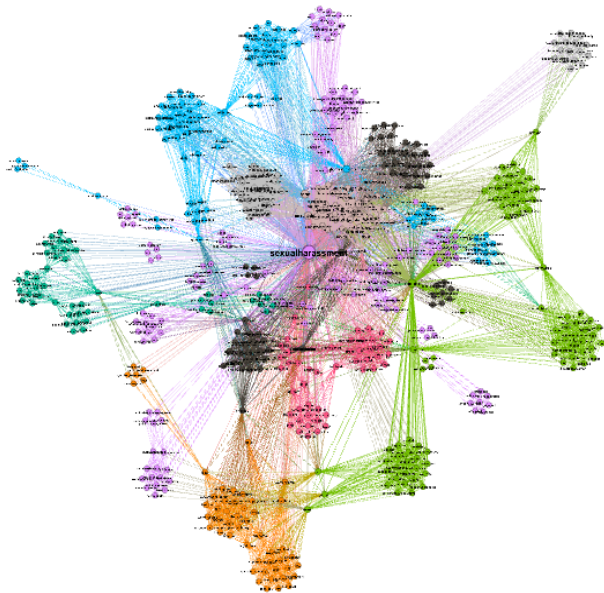


Fig. 2. Instagram - Full network on #sexualharassment

An edge is formed when two hashtags appear in one or more posts together, indicating co-occurrence. The weight of

an edge indicates the frequency in which the two hashtags co-occur together while the weight of a node refers to the frequency of the node appearing in the pool of data we crawled. We partitioned the data using modularity where higher modularity represents better quality of the partition. Since #sexualharassment is the main keyword and the overarching topic of this project, we performed the hashtag network analysis on this important keyword.

Fig. 2 shows the full network on the hashtag #sexualharassment, displaying 12 different communities with an overall score of 0.562 for Modularity. This means that the network has a rather moderate modularity where there are a few dense connections between nodes within the community but sparse connections between nodes in different communities [4]. The center of the network consists of the targeted hashtag node #sexualharassment. It is identified that this node connects with most of the other communities together.

Source	Target	Weight
sexualharassment	sexualabuse	34
sexualharassment	rapevictim	28
sexualharassment	sexualassault	27
sexualharassment	awareness	21
sexualharassment	violenceagainstwomen	21
sexualharassment	childsexualabuse	19
sexualharassment	cptsd	19
sexualharassment	manipulation	19
sexualharassment	metoo	19
sexualharassment	survivors	19

Fig. 3. Instagram - Top 10 co-occurring hashtags for #sexualharassment

Fig. 3 shows the top 10 co-occurring hashtags for #sexualharassment based on the descending order of the respective edge weights. With these insights, AWARE can use a combination of relevant hashtags for their public campaigns – in order to increase the reach of their social media posts.

### B. Twitter

Narrowing down to only sexual harassment related tweets, each node represents a user. An edge forms when one user likes or retweets the tweet, indicating the flow of information from one user to another. The edge weight indicates the number of unique post interactions between two users.

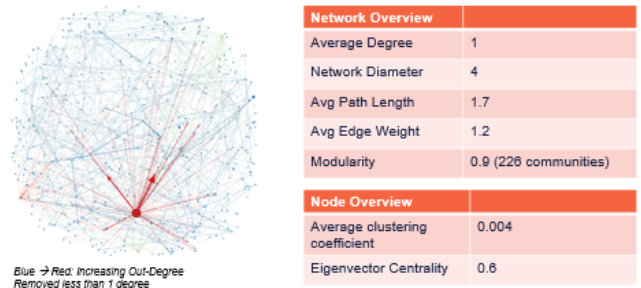


Fig. 4. Twitter - Removed nodes with degree range  $\leq 1$  (either in or out-degrees)

In community detection, we aim to detect communities of nodes where interactions are relatively frequent. Networks with high modularity have dense connections between nodes within the module but sparse connections between nodes in different

modules [4]. The network we derived from our Twitter dataset has a high modularity value of 0.9 (max modularity =1) and forms 226 communities. Given that the total number of nodes is 1,262, having 226 communities with five nodes indicates that there are no distinct ‘large’ communities in this network. This is further supported by a very low average clustering coefficient value of 0.004. This indicates that our Twitter network is unstable and imbalanced. From this, we inferred that the existing twitter network is not ideal for disseminating information as information may only reach certain clusters.

By ranking the nodes by descending out-degree centrality, we identified the top 10 users whose tweets are highly liked and retweeted. It is expected that AWARE and news platforms such STcom, BBC, CNA would top the charts since they are the main outlets that disseminates information. Interestingly, we identified local independent users within these top 10 users. Since these users are highly popular, AWARE can engage these users to spread information and increase the depth of their outreach efforts. These popular local salesmen are: 1) Kirsten Han (@kixees), a Singaporean journalist and activist, 2) Hugh (@vinsonhugh), Tattoo Apprentice, and 3) TheAsianFeminist (@theasianfmnst), a fan page that caters and shares news on women and feminism from 50 different countries in Asia.

No.	Id	Out-Degree	Followers Count
1	awarenews	297	3,998
2	STcom	239	1,036,800
3	BBCNews	155	10,637,923
4	kixees	123	12,332
5	ChannelNewsAsia	122	950,426
6	vinsonhugh	93	603
7	davidbrunelle	88	6,900
8	equalitynow	88	48,501
9	brianklaas	85	215,785
10	theasianfmnst	85	6,587

Fig. 5. Twitter - Popular salesmen based on descending out-degree

By ranking the nodes by descending betweenness centrality, we identified the top 10 users who are brokers in the network. Apart from news platforms, we identified local independent users within these top 10 users. Since the connected has a special gift in connecting disparate communities, they will be key in transmitting information in the existing highly disparate twitter social network. AWARE can engage these users to increase the breath of their outreach efforts. These vital local connectors are: 1) Bunnyricecakes (no description available), 2) Kixex (@kixees), 3) Visakan Veerasamy (@visakanv), an author, 4) Kellynn (@KellynnWee), a research associate at NUS Asia Research Institute, and 5) Thelocalrebel (@thelocalrebel), a local independent feminists collective.

#### IV. CONCLUSION & RECOMMENDATIONS

In this study, we performed digital social listening on conversations about sexual harassment. After speaking to

AWARE, we found out that AWARE aims to achieve 1000 pledges for AimForZero Campaign by Nov 2020. However, the pledge rate is becoming stagnant. As such, we are proposing to AWARE to adopt a three-pronged approach when pushing out more online awareness on the AimForZero Campaign.

Leveraging on the Law of the Few, we are proposing to AWARE to engage the potential salesmen and connectors identified in our analyses to increase the depth and breadth of their online outreach efforts. In addition, they can recruit the Supporters of AWARE as evangelists to promote the sexual harassment cause. Dealing with such a complex issue where there is ambiguity in the definitions for sexual harassment or assault, critical individuals like the translators are needed to overcome the problem of the chasm - these are individuals who take ideas and information from a highly specialized world and translate them into a language that the rest could understand or comprehend.

Since the existing network is an imbalanced network consisting of many small disparate communities, it is important to create content that can resonate within small social circles of friends [5]. Leveraging on the concept of homophily, we can invoke peer to peer influence. For example, among the popular salesmen, @kixees and @vinsonhugh’s profile personality are worlds apart and each has their unique social group. @kixees is probably able to connect well with the general public since she’s a journalist. While @vinsonhugh is probably able to connect well with young people and artists because they are similar in their interests and beliefs. As such, by creating content customized for various groups and spreading them through homophilic peers, the extent of influence will be more effective.

#### REFERENCES

- [1] Teng, A., & Ang, J. (2019). Most sexual misconduct cases on university campuses go unreported: Survey. The Straits Times. Retrieved from <https://www.straitstimes.com/singapore/education/most-sexual-misconduct-cases-on-university-campuses-go-unreported-survey>
- [2] Elangovan, N. (2019, November 3). Monica Baey, who sparked a national debate on sexual harassment, among women awarded for gender equality efforts. The Straits Times. Retrieved from <https://www.todayonline.com/singapore/nus-undergraduate-monica-baey-among-three-women-awarded-gender-equality-efforts>
- [3] Voices@SMU. (n.d.). Supporting a Harassment-Free Environment. Retrieved from Singapore Management University Supporting a Harassment-Free Environment Module
- [4] Ji, X., Machiraju, R., Ritter, A., & Yen, P. Y. (2015). Examining the distribution, modularity, and community structure in article networks for systematic reviews. In AMIA Annual Symposium Proceedings (Vol. 2015, p. 1927). American Medical Informatics Association.
- [5] Adams, P. (2011). Grouped: How small groups of friends are the key to influence on the social web. New Riders.