Xinchen Yu, Afra Mashhadi, Jeremy Boy, Rene Clausen Nielsen and Lingzi Hong (2022): Causal Impact Model to Evaluate the Diffusion Effect of Social Media Campaigns. In: Proceedings of the 20th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centered Computing on the Design of Cooperation Technologies - Exploratory Papers, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.48340/ecscw2022_ep10

Causal Impact Model to Evaluate the Diffusion Effect of Social Media Campaigns

Xinchen Yu[1], Afra Mashhadi[2], Jeremy Boy[3], Rene Clausen Nielsen[4], Lingzi Hong[1]

[1]University of North Texas
[2]University of Washington Bothell
[3]UNDP Accelerator Labs
[4]UN Global Pulse
xinchenyu@my.unt.edu
mashhadi@uw.edu
jeremy.boy@undp.org
rene@unglobalpulse.org
lingzi.hong@unt.edu

Abstract. Organized information campaigns on social media platforms have influence on collective opinions through processes such as social influence and majority opinion formation. Evaluating the effect of such campaigns has become a critical question. We proposed a method by first characterize user engagement and the semantics in public discussions with social media data, then apply a causal impact analysis to measure the effect. We conducted a case study to examine the effect of the 16 Days Campaign (a campaign organized by UN Women) through changes in public discussions of the MeToo, which is a related topic the campaign was aimed to impact. Results showed there were significantly more discussions in MeToo after the launch of the campaign. Hashtags on 16Days topics were used more and by more people. The proposed methods evaluate the direct and indirect diffusion effect of a campaign by quantifying the difference had the campaign not taken place based on social media data. The method enables to evaluate the overall outcome of collaborative work in a social media campaign.

Introduction

Social media campaigns refer to campaigns sustained by coordinated efforts to achieve specific goals with information spreading on social media platforms. Participants or volunteers are guided to post certain content under the agenda of

the campaign by using strategies such as spreading certain propaganda or diverting public attentions from important issues to shape collective opinions online (Bradshaw and Howard, 2018). Collective opinions may lead to individual behavioral changes, such as voting preferences (Aral and Eckles, 2019), or societal changes in policy making and legislation (Fileborn and Loney-Howes. 2019; Freischlag and Faria, 2018). How to evaluate the collaborative effort and the impact of social media campaigns has become a critical question. Campaigns oftentimes implement different strategies, for example, promoting through influencers or coalition with trending topics. The campaign managers hope to know whether the campaign achieved the expected outcomes by attracting public attentions or influencing people. Evaluating the effect enables to compare and identify effective campaigns. It may also help defense the impact of malicious campaigns that spread conspiracy theories or misinformation (Badawy et al., 2019).

Different methods have been implemented to investigate the effect of social media campaigns. Most of the evaluations are based on campaign participants, thus, to survey participants before and after campaigns (Thompson et al., 2020) or to compare the differences between campaign participants and controlled groups (Breza et al., 2021). However, many social media campaigns do not have either the controlled group setting or identifiable participants, not to mention that the collection of individual data such as survey responses or digital traces is difficult and costly (Aral and Eckles, 2019). In addition, social media campaigns are found to have spillover effect in many ways (Lee et al., 2018; Dincelli et al., 2016). Such effect cannot be measured with only data from campaign participants. Other studies examine outcomes of social media campaigns by analyzing the related tweets after certain events. However, tweets on the examined topics may exist before the campaign, the results may overestimate the effect of the campaign. The results are mostly descriptive with visualizations showing the temporal trend after the events. For example, Samuel et al. (2020) analyzed the emotional consequences of the public after the reopening policy during the COVID-19. Quantitative evaluation of the effect, thus, to measure the degrees of change had the event not happened is few. Recently, several studies proposed to statistically analyze the relation between social media campaigns and certain voting (Aral and Eckles, 2019) or protesting behaviors (Ahmed et al., 2017) with causal inference analysis. Few have applied the causal impact model to the evaluation of collective opinions impacted by social media campaigns. We apply the causal impact model to analyze the relative effect of a social media campaign, the 16Days campaign, on collective opinions in MeToo.

The study presented here has two main contributions. First, we design a set of metrics to characterize users' participation and the trending topics in collective opinions with social media data. The metrics can capture the diffusion effect of social media campaigns with data at the aggregated level, which doesn’t require

the identification of campaign participants or the use of individual level data. Second, we present a causal inference method for the quantitative evaluation of the diffusion effect of a social media campaign, which measures the difference between the observed values and the counterfactual values had the social media campaign not taken place. The causal impact model enables to measure the accumulated impact during a time range.

We applied the method on a case study and examined changes in the MeToo discussions after the launch of a social media campaign, the 16Days, which used hashtags such as #HearMeToo to raise discussions in MeToo and geared towards the awareness of women's rights and empowerment. Results showed that the 16Days campaign may have led to increased engagement in MeToo and more discussions on women empowerment topics, illustrating how a social media campaign may indeed influence public discussions.

The rest of this article is organized as follows: In Section 2, we review the related work on information diffusion on social networks and methods to evaluate the effect of social media campaigns. Section 3 presents on research methodology including an overview of the causal impact model, and the data we used for the case study. Section 4 presents our analysis and findings. Section 5 discusses the method and implications. Section 6 concludes this article and the limitations.

Related Work

Organized social media campaigns have been found to be implemented in many countries, which use strategies such as promoting certain content to cause it a trend, shaping discussions through comments, attacking opponents, and diverting conversations from important issues. Some of the campaigns could become a potential threat to the democratic society (Bradshaw and Howard, 2018).

Diffusion Effect in Social Networks

Several studies have investigated the diffusion of campaign information on social media. For example, Badawy et al. (2019) characterized the diffusion paths of messages from Russian Internet Research Agency's information campaigns during the 2016 US presidential election; Ferrara et al. (2020) analyzed the diffusion of automated misinformation in different user groups during the 2020 US presidential election. Recent studies have also proposed methods to identify the initial set of nodes (Smith et al., 2018) or interaction of nodes (Myers and Leskovec, 2012) that can maximize the diffusion of information in social networks.

However, the impact of social media campaigns is not only limited to the users involved in the diffusion of campaign information. A line of research investigates mechanisms how campaign information influence public opinions on social

media. The diffusion of campaign information can be amplified under the effects such as social affordance (Lee et al., 2018), homophily (Dincelli et al., 2016), and social contagion (Coviello et al., 2014). Homophily describes nodes sharing similar characteristics tend to act out similarly (Dincelli et al., 2016). Social contagion is used to characterize mimicry practice, such as the adoption or formation of opinions from social contacts due to network diffusion (Christakis and Fowler, 2013). The diffusion between agents is found to be moderated by many factors, for example, the sociocultural distance (DellaPosta et al., 2015; Yu et al., 2020 (a); Yu et al., 2020 (b)) and the interpreted meaning of the practice based on the distribution of all practices in the population (Goldberg and Stein, 2018). The initial exposure increases probability of diffusion, but the probability can be suppressed when the exposure comes to saturation (Hodas and Lerman, 2014). Yoo et al. (2019) built a self-exciting point process model to examine how the diffusion of information might be influenced by the parallel diffusion of similar content. Results showed that the diffusion effect can be inhibited or amplified depending on the network structure. These studies show the impact of a social media campaign is not only through the direct diffusion of campaign information but can be more intricate and complex. Therefore, the evaluation of campaign effect is partial if only campaign participants are considered, it requires an examination of the systematic outcome.

Evaluating the Effect of Campaigns

A lot of recent studies have paid attention to evaluating the effect of social media campaigns with a focus on campaign participants. These studies were conducted by comparing participants before and after the campaign (Thompson et al., 2020), people who are in campaign regions and are not (Buller et al., 2021), or people in regions with campaign activities of high and low intensity (Breza et al., 2021). For example, Thompson et al. (2020) evaluated the effect of a social media campaign on changing the mental health stigma in student participants. Breza et al. (2021) used behavioral data collected from Facebook to investigate how a social media campaign affected the mobility behaviors of people during COVID-19. The evaluation relies on the implementation of comparative experiments with controlled groups or is based on the survey data or digital traces from individuals. Such data may not be readily available in many cases. Most of the campaigns are not designed to be launched in certain geographical areas. It is also difficult to identify campaign participants and collect behavioral data from individuals (Aral and Eckles, 2019). In social media campaigns, people who are not campaign participants could be affected through relationships and the diffusion of information (Dincelli et al., 2016; Coviello et al., 2014). These methods cannot measure the spillover effect.

Aral and Eckles (2019) proposed to use causal inference to evaluate how social media manipulation may have influenced the US presidential election outcome.

Causal statistical analysis can be applied to analyze public opinion and behavior change across individuals and subpopulations by measuring deviations from expected behaviors due to manipulative factors. In this study, we propose to apply the causal impact method to evaluate the outcomes on collective opinions with social media data.

Social Media Data to Study Public Opinions

A lot of public discussions are happening on social media platforms. Social media data, therefore, become an important source of information to understand public opinions. Several studies have used social media data to analyze the consequences of events such as protests (De Choudhury et al., 2016, Ahmed et al., 2017; He et al., 2015), implementation of new policies (Samuel et al., 2020), public health incidents (Gaspar et al., 2016), and social media campaigns (Badawy et al., 2019). Samuel et al. (2020) focused on characterizing the public sentiment trends for the policy decision in COVID-19 pandemic. Hong et al. (2016) analyzed the trend of different topics after the Ferguson unrests. These studies generate descriptive results about the sentiment and topics of discussions and interpret the relations of the results to the events. Some studies used statistical analysis, for example, De Choudhury et al. (2016) unpacked the relations of thoughts, opinions, and sentiments in the discussions of the "Black Lives Matter" on Twitter and perspectives of offline protests. Ahmed et al. (2017) discovered a high resemblance in emotions of anger and anxiety between online discussions and offline campaigns. The studies are often used to modeling collective opinions echoing to offline events. Few have applied the causal impact method to analyze the relative effect on public opinions impacted by social media campaigns.

Research Methodology

This study aims to measure the impact of a social media campaign on the collective discussions on a related theme, identifying if the campaign leads to more public attention on the campaign agenda. On social media platforms such as Twitter, discussions on certain topics are symbolized by a set of hashtags. A campaign can interact with users using certain hashtags and expose the campaign claims to people sharing similar interests and impose influence. The diffusion of campaign information may lead to increased participation on certain topics and change public opinions through the process of social influence (campaign claims spread to people through social contacts) and homophily (users with similar interest show conformity in spreading information of the campaign) (Dincelli et al., 2016). Further, the campaign may bring public attention on specific claims and bring more discussions related to the claims (Edwards and Marullo, 1995). The dynamics of the diffusion process can be complex under the effects of

confounding factors (Monsted et al., 2017). We focus on the method to evaluate the outcomes of diffusion, considering that such interactions may lead to participation and content changes in collective opinions.

We conducted a case study on the effect the 16 Days of Activism (16Days) on public discussions of the MeToo. The 16Days is an annual international campaign sponsored by United Nations General Secretary and organized by UN Women that kicks off on 25 November 2018, the International Day for the Elimination of Violence against Women, and runs until 10 December, Human Rights Day. In 2018, the 16 Days campaign was designed to be associated with the MeToo movement online using hashtags such as #HearMeToo. The online campaign may trigger contagious ideas or behaviors in the discussions of MeToo. We aimed to evaluate the diffusion effect of the campaign and monitored changes in the MeToo. The campaign may have influenced public discussions in other topics that were not part of the MeToo. However, it is impossible to monitor all discussions on social media platforms.

Data Collection

Since the base of the campaign is in the US and most of the campaign conversations are in English, we used English tweets for the evaluation. With the help of UN Women specialists, hashtags of the MeToo movements and the 16Days campaign were identified (Table I). We used Crimson Hexagon ForSight, now Brandwatch Consumer Research (Crimson Hexagon, 2019), to collect two types of data: i) tweeting activities in the MeToo that exclude tweets with 16Days campaign hashtags. We collected time series data such as daily volumes of all tweets and unique tweets; and ii) sampled MeToo and 16Days tweets that are used for topic analysis. In total, there were 32,249,394 tweets posted in the MeToo movement from January 1, 2016, to January 15, 2019. We collected 313,451 sampled tweets, 151,398 of which are tweets of the 16Days campaign from 66,619 users in a 62 days' time span, 162,053 are tweets of the MeToo movement from 110,522 users in a 1,111 days' time span.

Table I. Hashtags and sampled tweets of the MeToo movement and the 16Days campaign.

	Hashtags	#Tweets	#Users	#Days
MeToo movement	#MeToo, #WithYou, #WeToo, #TimesUp, #TimeisNow	162,053	110,522	1,111
16Days campaign	#GBVteachIn, #GBV, #endGBV, #16daysofactivism, #endVAW, #endVAWG, #womensrights, #humanrights, #ILOendGBV, #StopGBVatWork, #RatifyILO190, #16days, #orangetheworld, #16DaysCampaign, #HearMeToo	151,398	66,619	62

Characterizing User Participation and Trending Topics in Time Series

We propose four metrics to measure the scale of user participation in MeToo: 1) total volumes of the posts, which reflects the scale of public discussions on the topic and partially reflected the occupation of public attention; 2) total volumes of unique posts, which measures the amount of new content created; 3) number of unique links in posts, which measures the diversity of exogenous information sources, such as news websites and videos; and 4) number of unique hashtags used, which reflects the diversity of information generated from the social media platform. All are computed in a certain temporal granularity. In the case study, we compute daily values.

Topical changes are measured based on the use of hashtags. Daily frequency is computed as number of posts with the hashtag. Considering the total volumes of tweets on MeToo could vary by day, daily proportion is used, which measures the proportion of tweets containing a certain hashtag to the total volume of MeToo posts. Active participants of a topic may advocate similar content for hundreds or even thousands of times (Shao et al., 2018). Therefore, we introduce daily weight in participants, which is the average ratio of a hashtag used by all users in MeToo. The metric is to capture the weight of a topic in the MeToo discussion. To further measure whether the hashtags with significant changes are relevant to the campaign, we compute the semantic similarity of the hashtag to the MeToo hashtag and campaign hashtag. The semantic embedding of these hashtags is learned by the global vectors for word representation (GloVe) algorithm. The GloVe embedding algorithm incorporates global statistics, i.e., word co-occurrence probabilities, and local context information, i.e., the neighboring words in a window, to obtain vector representations for words (Pennington et al., 2014). The similarity of two hashtags is measured by cosine similarity of the vector representations of hashtags.

Causal Impact Analysis Model

We conducted the causal impact analysis by applying the state-space model, which predicts the counterfactual time series indicating the status had no intervention taken place based on control groups (Brodersen et al., 2015). The model is chosen over other causal inference models as the effect is evaluated based on time series data. The model has two components: (1) the observed values capture the trend of the states and the disturbance in the observation. (2) the state is assumed to be a latent process with variances due to noises. The hidden or latent process is assumed to be a Markov process. The states generate observations, and these observations are only dependent on the states. The state is composed of local linear trend, seasonality, and contemporaneous covariates with static coefficients.

A structural model can be inferred based on the observed values in control groups (Brodersen et al., 2015). The observed values are related to the states, while the states are inferred from seasonal patterns, previous states, and potential noises. The model considers noisy signals, for example, users may discuss other issues using movement hashtags and contribute to the total volume of movement discussion. The model can be used to predict, given the unchanged dynamics of states and the existence of disturbance, what the observations could be. We can thus build a structural model based on historical observations and compare the observed values with the predicted values had the intervention not taken place.

The differences between the observed time series and the counterfactual time series reflect the impact of an intervention event. Time-series data usually have fluctuations. It is difficult to determine the impact level at a discrete time point. We measure the impact by accumulating all the differences in a time range and compute the relative effect in total. To make sure the effect size is comparable over time series of different measurements, we compute the relative effect as a percentage of summed differences to the summed counterfactual time series values. The causal impact analysis will be applied to the following measurements using a Python package Causal Impact (Fuks, 2020).

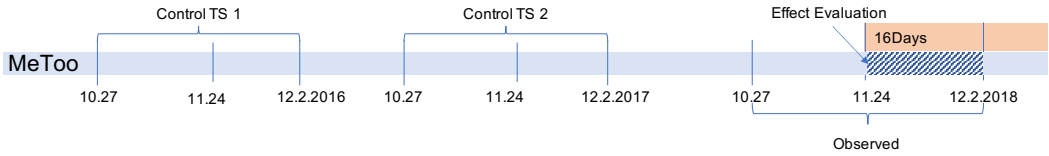

Figure 1. Control groups and effect evaluation time window in the timeline of MeToo.

Effect Evaluation

The 16Days campaign used the hashtag #HearMeToo to encourage people participate in the MeToo movement from Nov 24, 2018. The 16Days campaign can be seen as an event that cause variances in the observed time series of MeToo. The effect was evaluated by comparing the observed time series of MeToo with the synthetic time series (counterfactual), which were inferred based on historical data (control groups) assuming no interventions had happened.

The compared time series was set as 1 week from the time of the intervention with a granularity of one day. Correspondingly, the counterfactual data was about the same week that consisted of seven daily values. To infer the counterfactual time series, we set control groups with three components: i) the observed time series in 4 weeks before the intervention, which is Oct 27, 2018 to Nov 24, 2018 (part of Observed in Figure 1); ii) the observed time series one year ago, with 4 weeks before the intervention date and 1 week after, which is Oct 27, 2017 to Dec 2, 2017 (Control TS2 in Figure 1); iii) the observed time series two years ago, with the five weeks the same as ii), i.e. Oct 27, 2016 to Dec 2, 2016 (Control TS1

in Figure 1). These parameters are applied to evaluate time series behaviors in the scale of user participation and discourses.

In terms of topic analysis, we identified the hashtags that had significant changes due to the campaign intervention and computed the similarity of these hashtags to #HearMeToo and #MeToo, which are separately key hashtags of the 16Days campaign and the MeToo. The GloVe embeddings of hashtags were computed with the tweets of MeToo and 16Days in Control TS1, Control TS2 and Observed time ranges. After removing the repetitive tweets, we pre-processed the tweets by removing links, special characters, punctuation symbols, and entities that start with "@", transforming all tokens to lower cases and stemming words. After pre-processing, we got the vector representation of all the hashtags with a length of 50 and computed the cosine similarity of hashtags.

Finding

Figure 2 shows the contrast between counterfactual values and the observed volumes of tweets from October 27 to Dec 2 of 2018 for all tweets that include retweets (a) and unique tweets (b). Both figures show that the observed volumes are significantly larger than the model predicted values based on historical time-series (counterfactual). The increasing trend reached to the max in the first three days since the launch of the campaign. After that, the difference between observed values and the counterfactual values becomes smaller, which shows the effect of the campaign weakens.

Table II. Relative effect of causal impact analysis.

	In one week		In two weeks	
	relative effect	95% interval	relative effect	95% interval
Daily volume	38.44%	(7.71%, 68.62%)*	28.93%	(7.89%,50.18%)*
Daily volume of original tweets	34.02%	(11.32%, 56.15%)***	37.67%	(21.41%,55.62%)***
Proportion of original tweets	-4.85%	(-16.82%, 8.0%)	3.45%	(-6.74%,13.4%)
Number of unique links	1.8%	(-8.74%, 12.57%)	-10.98%	(-18.36%,-3.38%)***
Number of unique hashtags	13.16%	(-2.34%, 28.21%).	-6.02%	(-17.46%,5.35%)

Significance levels: 0 '***'; 0.001 '**'; 0.01 '*'; 0.05 '.'; 0.1 ' '

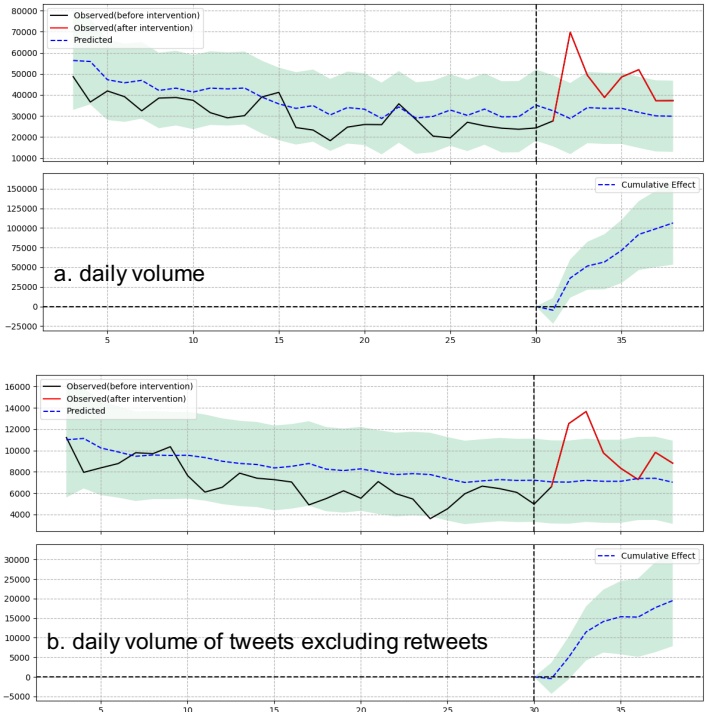

Figure 2. Time series of predicted and observed behaviors.

Table II shows the statistics of causal impact analysis. We found increased discussions in MeToo after the launch of the 16Days campaign, with more unique posts created by the participants and more retweeting. The observed volume of MeToo tweets reached to the maximum in the third and fourth day after the campaign was launched, then the effect was weakened afterwards. It is possible that the diffusion of campaign related information reached to the point of saturation, which inhibited the creation and diffusion of information (Hodas and Lerman, 2014).

Associated with the increasing volume, the unique external links contained in MeToo tweets did not change significantly but the unique hashtags increased. It reflects the increase of posts may come from the introducing of a new set of topics on Twitter rather than from the sharing of information from external sources. The causal impact model has considered potential noises and variances caused by random factors and attribute the effect of change to something happened at a certain time. In the case study, we set it as the day when the 16Days campaign was launched. The analysis of unique links and hashtags shows the increased participation of MeToo is probably not from other trending events, as the unique links indicating external sources do not change significantly in the first week. Rather, the increasing trend is more associated with information initiated from the platform, which is reflected from the more hashtags and more original content created in MeToo. To further reveal how the introduction of new

topics is related to the 16Days campaign, we conducted semantic analysis for online discussions.

Table III. Relative effect and the 95% interval of relative effect of hashtag usage. Top 10 hashtags with the largest relative effect are shown.

Hashtags	Weights in participants	Proportion of use	Frequency of use	Similarity	
				#MeToo	#HearMeToo
#freedom	25.11 (22.4, 27.81)***	45.93 (44.13, 47.61)***	39.14 (37.65, 40.89)***	0.44	0.32
#witchhunt	11.73 (10.6,12.92)***	-0.16 (-1.35,0.97)	-0.24 (-1.38,1.02)	0.08	0.08
#womensmarch	5.55 (4.5,6.55)***	3.98 (3.26,4.66)***	3.59 (2.86,4.3)***	0.35	0.39
#democracy	4.28 (2.41,5.93)***	26.35 (24.91,27.95)***	22.61 (21.05,24.32)***	0.33	0.20
#survivor	3.22 (2.31,4.13)***	1.3 (0.21,2.31)**	1.4 (0.34,2.38)***	0.47	0.37
#justice	2.95 (1.42,4.69)***	25.97 (24.34,27.69)***	24.07 (22.45,25.76)***	0.48	0.27
#metoomvmt	2.24 (1.26,3.18)***	1.36 (0.4,2.42)***	1.56 (0.59,2.54)***	0.31	0.33
#brexit	1.93 (0.92,3.06)***	0.28 (-0.83,1.36)	0.31 (-0.77,1.36)	0.15	0.27
#nomore	1.71 (0.29,3.22)**	0.12 (-1.12,1.38)	0.1 (-1.17,1.34)	0.25	0.21
#womenempowerment	1.63 (0.95,2.25)***	0.28 (-0.6,1.22)	0.2 (-0.77,1.1)	0.30	0.39

Significance levels: 0 '***'; 0.001 '**'; 0.01 '*'; 0.05 '.'; 0.1 ' '

For each hashtag that has been used in MeToo posts from 4 weeks before Nov 24 and 1 week after, we counted its daily frequency, daily proportion, and daily weight by users in TS1, TS2, and the observed time range. The values were set to zero if no hashtag was detected to be used in a day. We conducted a causal impact analysis with the time series of the three features for all hashtags. Table 3 lists the top 10 hashtags that show significant positive effect in daily weight by users in one week after the launch of the 16Days campaign, as we focused on topics that emerged due to the campaign. We also included the relative effect in terms of daily frequency and daily proportion as cross-references. We didn't present the 16Days hashtags here as they do not exist in previous MeToo tweets, thus all have significantly positive relative effect after the campaign.

There is a significant increase in the use of #freedom, which had a relative effect of 2511% in daily weight in users. The effect on daily frequency and daily

proportion is even larger. #Freedom was used in different contexts, for example, to call for a freedom culture from traditions – "... go out against #patriarchy, #misogyny and other abuse of women...#freedom #equality #secular", or to talk about the freedom of sex victims of human trafficking – "Let's end slavery... #freedom #slavery #escort #childabuse #metoo #stopthedemand...". The hashtag has a relatively higher similarity with #MeToo but is also seen in many cases that have gender perspectives. For example, "@XXX on #MeToo #GenderEquity #Freedom #quotes Gender justice has never been only about women, but about modernity, an expansion of democracy, ..."

Of the 10 hashtags, #womensmarch, #metoomvmt, and #womenempowerment have higher similarity with #HearMeToo compared to #MeToo. There is an estimated increase of 555% in #womensmarch, 224% in #metoomvmt, and 163% in #womenempowerment in terms of the daily weight in users. Both #womensmarch and #womenempowerment have an emphasis on the gender perspectives, which have overlaps with the claims of the 16Days campaign. The hashtag #womensmarch originated from a protest anti-women and offensive statements by politicians in 2017. The movement is mainly led by women and has a critical goal to end gender-based violence. Women's March is scheduled every year around January 20th from 2017. The variances cannot be a post-event effect as there are ten months' lag. The significantly increasing weight of #womensmarch could be partially attributed to the campaign. Hashtag #womenempowerment is mainly used in posts on gender equality. It calls for the inclusion of women in the decision-making process in economy and politics through "education, raising awareness, literacy and training" (Bayeh, 2016). The trending of these hashtags indicates more discussions on the gender-based claims and gender equality, which is what the 16Days advocates. It implies that the diffusion of campaign information may have lead participants in the MeToo movement to have increased attention on topics advocated by the campaign. Meanwhile, there is increased use of hashtags such as #freedom, #survivor, #justice, and #nomore, which although have higher similarity with #MeToo but are used in the 16Days claims that calls for social justice to sex abuse survivors. Besides, we observed the increased weight of #witchhunt, which is not quite related to the MeToo movement or the 16Days campaign, and #democracy and #brexit, which are mainly used in political discussions. The findings show that not all the increased discussions were caused by the diffusion effect of the campaign.

Discussion

To summarize the findings, we found that novel information was generated and circulated in the MeToo after the launch of the 16Days campaign even after removing the tweets that are directly brought by participants of the campaign. The increasing trend is associated with significantly more hashtags rather than the

number of external links, which could indicate the effect mainly come from online events. Further, the semantic analysis shows the weight of some topics (hashtags) has significantly increased, of which there are #womensmarch, #womenempowerment and #metoomvmt that have a higher association with the 16Days campaign. The results illustrate other external factors than the campaign, may together exert confounding impact on MeToo discussions.

In the case study, we evaluated the impact with public opinions on MeToo due to the relation between the campaign and MeToo discussions. To apply the method on other effect evaluation cases, there are several premises to make sure the results are relatable to the intervention of the event. The monitored activities should be an expected outcome brought by the campaign. The method helps to find out how much the monitored activities could be affected. Social media discussions could be subjected to external factors such as posts of influential figures or news. If there are identifiable factors, the effect of these factors should be considered in the interpretation of results.

The choice of controlled time series for training the model to generate counter factual values is important. In the presented case, we chose controlled time series from two years. The choice is partially due to the lack of data before 2016 for model training, as there are too few MeToo discussions before 2016. Each year we used activities in four weeks as previous states to infer the following week at the same time of the year. This allows to capture seasonable patterns. A longer time range for training may include large variations caused by significant events in previous years, while a shorter time range doesn't include enough data points to train the model. In other cases, effect evaluation analysis may consider similar reasons for the choice of time series to build the model.

We combined different perspectives of analysis to infer the impact of the campaign. It is highly probable the outcome doesn't attribute to one single factor. In fact, due to the existence of confounding factors, the identified relative effect is a maximum boundary for reference. We may use topic analysis to further investigate the relations. The method provides a complementary analysis that focused on the deviation from normal patterns had the event not taken place. It may not apply to new activities caused by the event, for example, the use of hashtags created by the campaign was not used before. The method is especially valuable to evaluate the differences from normal patterns caused by a campaign.

Finally, in the case study we mainly analyzed the use of hashtags. Combing with other types of content analysis, we may generate metrics that characterize different dimensions of public opinions. For example, we may generate time series of sentiment polarity to investigate if the campaign has significant emotional consequences in the public. By using automatic classification algorithms, we may identify tweets that belong to the misinformation categories. It will enable to identify whether a certain event or campaign led to significant increase or decrease of misinformation spreading. In this case, we mainly

explored the measurements that are valuable to evaluate the effect of the 16Days campaign.

Implications

Social media campaigns are collaborative work by campaign participants but could have impact on users who are not directly engaged in the campaign through diffusion effects or public attentions. Therefore, effect evaluation of campaigns should consider the overall outcomes rather than focus on the effects on participants. The proposed method enables to quantitatively evaluate the differences that can be induced by an event. Different from previous studies that focus on interpretating descriptive statistics of the outcomes after an event, the method considers the activities that could have existed due to previous states and seasonal patterns. It provides statistics indicating if the relative effect is significant due to the events or if the variances are generated by random noises.

The results presented have practical implications for campaign managers and activist organizations. First, our results echo the findings that an organization-driven campaign could be highly influential and impactful in shaping online discussions. In the case of MeToo, a movement that is often associated with celebrities, our results show that a timely and unified intervention by UN Women helped to spread the MeToo movement and broadening the movement through diversifying. Second such diversification could also lead to a more diverse set of audience and reach well beyond the celebrity focus. The method proposed in this study can be applied to the effect evaluation of other social media campaigns and allow for dynamic monitoring of the intervention by the campaign managers.

Conclusion

This study investigates the diffusion effect of a campaign on public discussions. We have proposed measurements to characterize movement participation from the perspectives of user engagement and discourses with social media data. We have applied a causal impact analysis to measure the relative effect due to the intervention of the campaign. The method is helpful to find the maximum boundary of the relative effect on collective opinions. Combining semantic analysis, it further reveals how the campaign influenced the discussions and other potential factors that might have affect the discussion.

This study is an initial exploration with several limitations. First, we have tried to capture the evolving of the collective opinions from different perspectives but there are other ways to characterize the participation and topics we did not include, for example, the number of individual participants or mentioned users. The analysis focused on the metrics that we could obtain through the summary statistics from the API (Brandwatch). That's the case for data-driven studies that

we can only develop methods and reveal insights based on the available data. However, it is notable that the causal impact analysis method is applicable to different time series data at different temporal granularity for evaluating the effect. Given more data, we will be able to present a more comprehensive evaluation. Second, the causal impact model we used has limitations. It is used to measure the differences had something not happened in a certain time point. As we mentioned, public opinions could be potentially affected by other factors that happen near the time point. It is possible there were chain reactions, for example, the increased public attention led politicians to involve in the discussions and bring new topics. However, the current state-space model has limitations that it cannot model the unidentified external factors or the effect of a series of external factors. Third, the case study presented here investigates the campaign in 2018, which was designed to be related to the MeToo movement. It would be a future work to obtain data from other years to compare the effect and identify effective strategies.

Acknowledgments

The authors would like to thank UN Global Pulse Lab and UNWomen for their time and support to shape this work.

References

Ahmed, S., Jaidka, K. & Cho, J. (2017), 'Tweeting india's nirbhaya protest: a study of emotional dynamics in an online social movement', *Social Movement Studies* 16(4), 447–465.

Aral, S. & Eckles, D. (2019), 'Protecting elections from social media manipulation', *Science* 365(6456), 858–861.

Badawy, A., Addawood, A., Lerman, K. & Ferrara, E. (2019), 'Characterizing the 2016 russian ira influence campaign', *Social Network Analysis and Mining* 9(1), 1–11.

Bayeh, E. (2016), 'The role of empowering women and achieving gender equality to the sustainable development of ethiopia', *Pacific Science Review B: Humanities and Social Sciences* 2(1), 37–42.

Bradshaw, S. & Howard, P. N. (2018), 'Challenging truth and trust: A global inventory of organized social media manipulation', *The Computational Propaganda Project* 1, 1–26.

Breza, E., Stanford, F. C., Alsan, M., Alsan, B., Banerjee, A., Chandrasekhar, A. G., Eichmeyer, S., Glushko, T., Goldsmith-Pinkham, P., Holland, K. et al. (2021), 'Effects of a large-scale social media advertising campaign on holiday travel and covid-19 infections: a cluster randomized controlled trial', *Nature medicine* pp. 1–7.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., Scott, S. L. et al. (2015), 'Inferring causal impact using bayesian structural time-series models', *The Annals of Applied Statistics* 9(1), 247–274.

Buller, D. B., Pagoto, S., Baker, K., Walkosz, B. J., Hillhouse, J., Henry, K. L., Berteletti, J. & Bibeau, J. (2021), 'Results of a social media campaign to prevent indoor tanning by teens: A randomized controlled trial', *Preventive medicine reports* 22, 101382.

Christakis, N. A. & Fowler, J. H. (2013), 'Social contagion theory: examining dynamic social networks and human behavior', *Statistics in medicine* 32(4), 556– 577.

Coviello, L., Sohn, Y., Kramer, A. D., Marlow, C., Franceschetti, M., Christakis, N. A. & Fowler, J. H. (2014), 'Detecting emotional contagion in massive social networks', *PloS one* 9(3), e90315.

Crimson Hexagon (2019). Last accessed 27 December 2019. URL: *https://forsight.crimsonhexagon.com/*

De Choudhury, M., Jhaver, S., Sugar, B. & Weber, I. (2016), Social media participation in an activist movement for racial equality, *in* 'Tenth International AAAI Conference on Web and Social Media', AAAI.
DellaPosta, D., Shi, Y. & Macy, M. (2015), 'Why do liberals drink lattes?', *American Journal of Sociology* 120(5), 1473–1511.
Dincelli, E., Hong, Y. & DePaula, N. (2016), 'Information diffusion and opinion change during the gezi park protests: Homophily or social influence?', *Proceedings of the Association for Information Science and Technology* 53(1), 1–5.
Edwards, B. & Marullo, S. (1995), 'Organizational mortality in a declining social movement: The demise of peace movement organizations in the end of the cold war era', *American Sociological Review* pp. 908–927.
Ferrara, E., Chang, H., Chen, E., Muric, G. & Patel, J. (2020), 'Characterizing social media manipulation in the 2020 us presidential election', *First Monday* 25, 11.
Fileborn, B., & Loney-Howes, R. (Eds.). (2019). *# MeToo and the politics of social change*. Springer Nature.
Freischlag, J. A., & Faria, P. (2018). It is time for women (and men) to be brave: a consequence of the# MeToo movement. *Jama* 319(17), 1761-1762.
Fuks, W. (2020). Pycausalimpact [Source code]. Retrieved from https://pypi.org/project/pycausalimpact/
Gaspar, R., Pedro, C., Panagiotopoulos, P., & Seibt, B. (2016). Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior*, 56, 179-191.
Goldberg, A. & Stein, S. K. (2018), 'Beyond social contagion: Associative diffusion and the emergence of cultural variation', *American Sociological Review* 83(5), 897–932.
He, J., Hong, L., Frias-Martinez, V. & Torrens, P. (2015), Uncovering social media reaction pattern to protest events: a spatiotemporal dynamics perspective of ferguson unrest, *in* 'International conference on social informatics', Springer, pp. 67–81.
Hong, L., Yang, W., Resnik, P., & Frias-Martinez, V. (2016, November). Uncovering topic dynamics of social media and news: the case of Ferguson. In *International Conference on Social Informatics* (pp. 240-256). Springer, Cham.
Hodas, N. O. & Lerman, K. (2014), 'The simple rules of social contagion', *Scientific reports* 4, 4343.
Lee, S., Ha, T., Lee, D. & Kim, J. H. (2018), 'Understanding the majority opinion formation process in online environments: an exploratory approach to facebook', *Information Processing & Management* 54(6), 1115–1128.
Leung, R. & Williams, R. (2019), '# metoo and intersectionality: An examination of the# metoo movement through the r. kelly scandal', *Journal of Communication Inquiry* 43(4), 349–371.
Mønsted, B., Sapiezynski, P., Ferrara, E. & Lehmann, S. (2017), 'Evidence of complex contagion of information in social media: An experiment using twitter bots', *PloS one* 12(9), e0184148.
Myers, S. A., & Leskovec, J. (2012, December). Clash of the contagions: Cooperation and competition in information diffusion. In *2012 IEEE 12th international conference on data mining* (pp. 539-548). IEEE.
Pennington, J., Socher, R. & Manning, C. D. (2014), Glove: Global vectors for word representation, *in* 'Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)', Association for Computational Linguistics, pp. 1532–1543.
Samuel, J., Rahman, M. M., Ali, G. M. N., Samuel, Y., Pelaez, A., Chong, P. H. J., & Yakubov, M. (2020). Feeling Positive About Reopening? New Normal Scenarios From COVID-19 US Reopen Sentiment Analytics. *IEEE Access* 8, 142173-142190.
Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.C., Flammini, A. & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications* 9(1), 1–9.
Smith, S. T., Kao, E. K., Shah, D. C., Simek, O., & Rubin, D. B. (2018, June). Influence estimation on social media networks using causal inference. In *2018 IEEE Statistical Signal Processing Workshop (SSP)* (pp. 328-332). IEEE.
Thompson, A., Hollis, S., Herman, K. C., Reinke, W. M., Hawley, K., & Magee, S. (2020). Evaluation of a social media campaign on youth mental health stigma and help-seeking. *School psychology review* 50(1), 36-41.
Yoo, E., Gu, B., & Rabinovich, E. (2019). Diffusion on social media platforms: A point process model for interaction among similar content. *Journal of Management Information Systems* 36(4), 1105-1141.

Yu, X., Daida, S. R., Boy, J., & Hong, L. (2020, October). The Effect of Structural Affinity on the Diffusion of a Transnational Online Movement: The Case of# MeToo. In *International Conference on Social Informatics* (pp. 447-460). Springer, Cham.
Yu, X., Daida, S. R., Bentula, L., & Hong, L. (2020). Characteristics of information spreading across nations. *Proceedings of the Association for Information Science and Technology* 57(1), e309.