2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

On the Causal Relation between Users' Real-World Activities and their Affective Processes

Seyed Amin Mirlohi Falavarjani*, Ebrahim Bagheri†, Ssu Yu Zoe Chou†, Jelena Jovanovic‡ and Ali A. Ghorbani*
*University of New Brunswick, Canada
†Ryerson University, Canada
‡University of Belgrade, Serbia
smirlohi@unb.ca, bagheri@ryerson.ca, ssu.chou@ryerson.ca, jeljov@fon.rs, ghorbani@unb.ca

***Abstract*—Research in social network analytics has already extensively explored how engagement on online social networks can lead to observable effects on users' real-world behavior (e.g., changing exercising patterns or dietary habits), and their psychological states. The objective of our work in this paper is to investigate the flip-side and examine whether engaging in or disengaging from real-world activities would reflect itself in users' *affective processes* such as anger, anxiety, and sadness, as expressed in users' posts on online social media. We have collected data from Foursquare and Twitter and found that engaging in or disengaging from a real-world activity, such as frequenting at bars or stopping going to a gym, have direct impact on the users' affective processes. In particular, we report that engaging in a routine real-world activity leads to expressing less emotional content online, whereas the reverse is observed when users abandon a regular real-world activity.**

***Index Terms*—Quasi-experiments, Causality, Affective Processes, Twitter**

I. Introduction

Online social networks play an increasingly important role in different aspects of people's lives and have attracted researchers to explore the possibility of understanding people's behavior and decision making mechanisms based on the content exchanged on online social networks [1]. While randomized controlled trials are quite expensive and cumbersome to run on online social networks, alternative causal methods such as *natural experiments* and *quasi-experiment designs* can be adopted to explore cause-effect relations between phenomenon observed on online social networks and user behavior. Such study designs have already yielded interesting findings in areas such as user dietary behavior [2] and suicide ideation [3]. These forms of causal studies have also explored the impact of users' online behavior on their offline activities. For instance, Stuck et al. [4] reported that Fitbit users who were active on the Fitbit social network and suffered from diabetes were six-times more likely to increase their daily steps for each additional social connection they had. Similarly, Althoff et al. [5] found that each addition of a new social connection on the Argus app increased the number of steps a user took per day by 400 steps. Such studies have established that, under certain conditions, online behavior can shape a person's offline activities.

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08...$15.00
https://doi.org/10.1145/3341161.3342918

However, fewer work investigated the impact of real-world activities on users' behavior on online social networks. For instance, Branley and Covey [6] are among the few to explore this and find that risky offline behaviours, such as excessive alcohol consumption, illegal drug use, self-harm, and dangerous pranks, were linked to increased exposure to risky online content. Kiciman et al. [7] reported that the content shared on online social platforms by alcohol consuming college students was different from other groups of students, and such difference could be used to predict academic success. Similarly, our work focuses on understanding the impact of users' offline real-world activities on their *affective processes* as expressed in the content shared online.

In this paper, we are specifically interested in answering two complementary research questions (RQs), i.e., whether engaging in (RQ1) or disengaging from (RQ2) regular real-world activities, such as frequenting a bar or going to a gym, has any causal impact on users' online affective processes such as anger, anxiety, and sadness, among others. To this end, we adopt a *quasi-experimental design* and systematically compare the affective processes of users who engage or disengage with real-world activities with that of two matched control groups. We find that both engaging in and disengaging from real world activities have a statistically significant yet inverse causal impact on users' affective processes. This is an important finding in that it can help in deciding on the best set of offline real-world activities for those users who express symptoms such as anxiety, anger, and sadness in their online social content.

II. Research Methodology

The research methodology used in our work is based on a *quasi-experimental design* [8]. Specifically, in our case,

TABLE I
FOURSQUARE CATEGORIES CONSIDERED FOR EACH VENUE CATEGORY. * IS THE WILDCARD SYMBOL AND (*) IS A CONTROLLED WILDCARD WHERE WE SELECTIVELY CHOOSE ONLY SPORTS RELATED KEYWORDS FROM THE LIST OF CATEGORIES.

Venue Category	Count	Foursquare Categories
Shopping	33	*Mall, *Store, *Shop, *Market, *Boutique
Bar	22	*Bar, *Pub
Restaurant	131	Restaurant, Buffet, Noodle House, Taco Place, Pizza Place, Steakhouse, BBQ Joint, Burger Joint
Sport	25	*Gym, (*)Studio, (*)Rink, (*)Center, Pool, (*)Club, (*)Court, (*)Field, (*)Ground

the quasi-experimental design took the form of an interrupted time series design with a *treated* and a *control* group. Given the goal was to discover the effect of users' offline activities on their online affective processes, the *treatment* in our experiment was engaging in or disengaging from an offline activity. The measured *outcome* of the applied treatment was its impact on users' affective processes. Accordingly, the treated group consisted of users who embarked on or abandoned an offline activity after a certain period of time, whereas the control group included those users for whom such engagement/disengagement with the offline activity was not observed. Briefly, our experiment consisted of four major elements explained below: (1) the dataset used for conducting the experiment, (2) the treatment and the treated group, (3) the control group(s), and (4) the outcome.

A. The Dataset

Considering the objective of our study, our dataset needed to include data on users' offline activities and their online behavior. We collected data from Twitter to represent users online behavior and from Foursquare to capture their offline activities. In order to link users across the two platforms, we utilized the Swarm application. This application allows users to share, on Twitter and Facebook, the places they visit based on Foursquare venues. This made it possible for us to identify users who were active on both Twitter and Foursquare and to systematically link their offline and online activities.

To curate the dataset, we gathered data from October 2014 to April 2017 using the Twitter API. To identify users with Foursquare check-ins, among the collected tweets, we identified those with Swarm information. Among the users who were on Twitter and shared their location using Swarm, we only retained those who had Swarm check-ins in at least 10% and at most 50% of their tweets to ensure (1) we had enough check-in information for each user and (2) that each user posted other tweets beside Swarm check-ins. We eliminated users who tweeted in languages other than English. This resulted in 17,220 users who were present on Twitter and frequently posted location information using Swarm.

The venues where the users had checked in were then pooled using Foursquare API and categorized into four venue categories as shown in Table I. The fact that our dataset was collected over a three-year time period, excludes any bias towards certain topics or sporadic user behavior. The dataset consists of 17,220 users, 6,523,257 check-ins at 211 Foursquare venue sub-categories, and 48,672,327 tweets. Hence, in terms of size, our dataset is comparable to the datasets used in similar studies, such as the dataset described in [4] which was collected from a commercially closed platform and the Twitter dataset used in [9], among others. It should be noted that all personally identifiable information were anonymized during data collection.

B. Treatment and Treated Group

The treatment is defined as participation in or abandoning of an offline activity. We define the treated group as those users who can be identified as impacted by the treatment based on their Foursquare check-ins. We study two treatments, namely (1) embarking on (going) and (2) abandoning (leaving) an offline activity. Each of these two treatments are studied in the context of four venue categories (see Table I). We perform our observations on each user over a four-month time period, which consists of two months prior to the treatment and two months post treatment. In order to form the treated group for each venue category, for the treatment of embarking on an offline activity (e.g., started frequenting bars), we identify and select those users who did not have any check-ins in the related venue category for a period of two months and subsequently made at least sixteen evenly distributed check-ins in the subsequent two months (i.e., at least one check-in every week, to assure that they checked in at relevant venues consistently during the two month period). So, for instance, the treated group for the activity of going to Bars would consist of those users who, based on their Foursquare check-ins, had not been to any bar venues for two months and subsequently visited such venues at least once a week for two months. A similar approach is adopted for abandoning (leaving) treatments. In these treatments, we ensure that the subjects in the treated group had consistently checked in at venues of the given venue category type at least sixteen times and at least once every week over a two-month period after which the subject abandoned venues in that venue category and did not make any check-ins in such venues for a two-month period.

C. Control Groups

We used the following two different approaches to form two distinct control groups:
Control Group 1: We performed Propensity Score Matching (PSM) [10] in order to rule out the impact of the confounding variables (Table II) that, if not handled, could potentially yield unreliable causal relations. In particular, the PSM method allows for balancing the treated and control groups such that

TABLE II
BALANCING TREATED AND CONTROL GROUPS USING PSM.

Variable	SMD
Number of Tweets	0.1947
Number of Twitter Followers	0.0327
Number of Twitter Friends	0.0916
Number of Check-ins	0.1903
Gender	0.12
Median Absolute SMD	0.12

the distributions of the measured confounding variables are similar in the treated and the control sub-populations [11]. Following the PSM method, users in the treated and control groups were matched based on their propensity scores. The propensity scores were obtained from a logistic regression model fitted to predict the probability of a user being assigned to the treated group given the set of observed confounding variables of Table II. In Table II, we report the standardized mean difference (SMD) of the confounding variables in order to show that the users in the treated and control groups are well balanced. It is generally suggested that the absolute SMD value should stay below 0.25 [12]. Given all the observed variables in Table II are in the acceptable range, the experiment can be considered random and hence, as mentioned by Althoff et al. [5], other unobserved variables are expected to be well balanced and would not confound the findings.

Control Group 2: When using the PSM method, it would be ideal to control for all potentially confounding variables such as *age* and *geographical location*. However, such information is neither reliably nor comprehensively available on Twitter or Foursquare. In order to eliminate all demographic confounding variables, we created the second control group by matching every user in the treated group in the treated year with herself but in a non-treated year. By doing so, we are able to eliminate all demographic confounding variables as the matched users in both groups are exactly the same. This also eliminates the impact of time-of-year, since the outcome of both groups are for the same time of the year (but in different years).

The employment of two control groups, one based on propensity score matching and the other based on matching the user with herself but in a non-treated time period, allows us to generalize our findings and eliminate any confounding variables. If the same degree of causality is observed based on both control groups, then the causal conclusion can be accepted with a high confidence given that the confounding variables have been properly controlled.

D. Outcome

We measure users' affective processes based on LIWC [13], a widely-used text analysis program that computes the counts of words in different categories such as various linguistic dimensions, grammatical structures, and social and psychological processes. To measure users' affective processes, we used six related dimensions from the LIWC's *psychological processes* category namely, affect, anger, sadness, anxiety, positive emotion, and negative emotion. We ran each tweet through LIWC to obtain a measure for each of the six affective sub-processes.

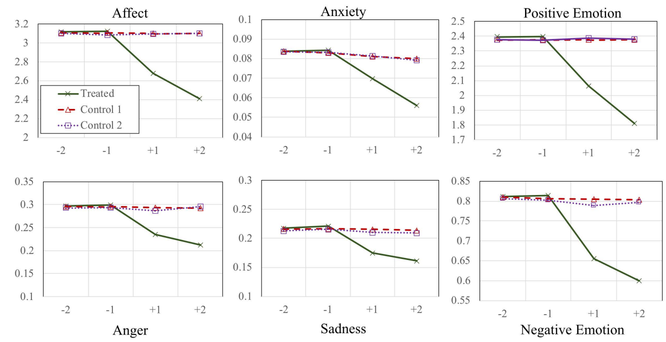

Fig. 1. Impact of engaging in an offline activity. The outcome for the treated and control populations over different *affective processes* is shown on the y-axis. Months -1 and -2 refer to the behavior prior to the treatment and Months +1 and +2 are two months after the treatment. The differences between treated and control groups in Months +1 and +2 are statistically significant based on a paired t-test with $p - value < 0.0001$ while no significant differences is observed in Months -1 and -2.

Let u be a user in one of the treated or control groups who posted m tweets in Month i, we measure the outcome as:

$$o_\theta^i(u) = \frac{\sum_{t \in m} \theta(t)}{|m|} \tag{1}$$

where $\theta = \{$affect, anger, sadness, anxiety, positive emotion, and negative emotion$\}$ is one of the six LIWC affective sub-processes and $\theta(t)$ is the outcome of sub-process θ on tweet t. For example, $o^1_{\text{anxiety}}(u)$ would calculate the average LIWC anxiety sub-process score over the tweets of user u in the first month whereas $o^2_{\text{anger}}(u)$ would measure the average LIWC anger sub-process score of user u's tweets in the second month.

We measure $o_\theta^i(u)$ based on the six LIWC sub-processes for the users in the treated group as well as their two corresponding control groups for the four-month time period, i.e., two months prior to the treatment (Months -1 and -2) and two months after the treatment (Months +1 and +2).

III. FINDINGS

The main research objective of our work is to study whether a causal relation can be found between users' offline activities and the online representation of their affective processes. This research objective led to the following two research questions:

RQ1. Whether *engaging in* a new offline activity impacts a users' online affective processes?

RQ2. Whether *disengaging from* a routine offline activity affects a users' online affective processes?

In order to explore these two research questions, we formed a treated group and two corresponding control groups for each of the venue categories introduced in Table I once for the case when the treatment is to engage in activities related to venues in that category and once for when the treatment is to disengage from the activities related to that category. For

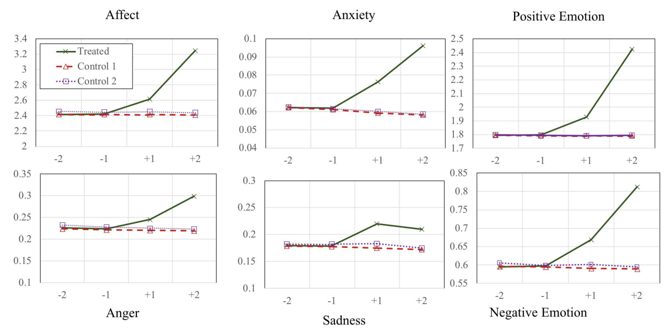

Fig. 2. Impact of disengaging from a regular offline activity. Figure structure is similar to Figure 1.

instance, we form a treated group and its two control groups when the treatment is to frequently visit bar-related venues and also a separate treated group and its related control groups for the situation when the treatment is to stop going to bar-related venues.

Figure 1 summarizes our findings with respect to RQ1. As shown in the figure, the treated group fully aligns with both control groups prior to the treatment in Months -1 and -2. However, after the treatment, the treated group deviates from the two control groups while both control groups remain fully aligned. Based on this observation, we find that engaging in a new offline activity leads to a significant causal impact on all six affective sub-processes of the treated users. Our results show that once a user embarks on and remains engaged with a new real-world activity, the scores of all six affective sub-processes reduce significantly. This further means that the users who become actively engaged in a real-world offline activity expose less affective content in their social media posts. In other words, such users employ less emotions in the contents they post. Furthermore, this reduction in affective content is true for both positive and negative content alike. As such our response to RQ1 is that active engagement in real-world offline activities causally impacts users' affective processes in such a way that the users' content becomes less expressive of their affective states.

Figure 2 depicts our findings related to RQ2, that is, how abandoning a frequent offline activity can impact users' affective processes. As shown on the figure and similar to our observations regarding RQ1, disengaging from a routine offline activity significantly impacts users' affective processes. However, in this case, we observe that disengaging from a real-world offline activity positively impacts users' affective processes, which means that a higher degree of emotional content, be it positive or negative, is shared on the social platform. Therefore, our finding related to RQ2 is that disengaging from a frequent real-world offline activity causally impacts users' affective processes and leads to users' more emotional expression on the online platform.

It is important to notice that in both Figures 1 and 2, the difference between the behavior of the treated group and both of the control groups is negligible and statistically insignificant in Months -2 and -1. However, the difference is statistically significant compared to both control groups in Months +1 and +2, while the behavior of the two control groups remains quite similar without any noticeable changes. Therefore, considering that with the two control groups we have controlled for most potentially relevant confounding factors and still observed statistically significant differences between the treated group and each of the control groups, we can conclude that the observed statistical differences are primarily due to the treatment and as such the conclusions can be considered to be reliable.

IV. Concluding Remarks

We have presented an observational study, based on a quasi experimental design, that examined whether users' real-world activities impact the way they express their emotional and affective states. By carefully controlling confounding variables through propensity score matching and a novel way of forming a control group, we have shown that engaging in or disengaging from a regular routine offline activity leads to meaningful changes in users' affective processes. The work in this paper opens up interesting avenues for future work: given the observed causal impact, we are now interested in (1) exploring similar causal relations to other psychological processes including social, cognitive, perceptual, and biological processes, and, (2) building recommendation models that would suggest the best offline activities for positively affecting users' specific psychological processes.

References

[1] J. M. Hofman, A. Sharma, and D. J. Watts, "Prediction and explanation in social systems," *Science*, vol. 355, no. 6324, pp. 486–488, 2017.
[2] Y. Mejova, H. Haddadi, A. Noulas, and I. Weber, "# foodporn: Obesity patterns in culinary interactions," in *Digital Health*, 2015, pp. 51–58.
[3] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media," in *CHI*, 2016, pp. 2098–2110.
[4] D. Stück, H. T. Hallgrímsson, G. Ver Steeg, A. Epasto, and L. Foschini, "The spread of physical activity through social networks," in *WWW Conference*, 2017, pp. 519–528.
[5] T. Althoff, P. Jindal, and J. Leskovec, "Online actions with offline impact: How online social networks influence online and offline user behavior," in *WSDM*, 2017, pp. 537–546.
[6] D. B. Branley and J. Covey, "Is exposure to online content depicting risky behavior related to viewers' own risky behavior offline?" *Comput Human Behav*, vol. 75, pp. 283–287, 2017.
[7] E. Kiciman, S. Counts, and M. Gasser, "Using longitudinal social media analysis to understand the effects of early college alcohol use," in *ICWSM*, 2018, pp. 171–180.
[8] J. DiNardo, "Natural experiments and quasi-natural experiments," *The new Palgrave dictionary of economics*, pp. 1–12, 2016.
[9] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "The effect of collective attention on controversial debates on social media," in *Web Science*, 2017, pp. 43–52.
[10] M. Caliendo and S. Kopeinig, "Some practical guidance for the implementation of propensity score matching," *Journal of economic surveys*, vol. 22, no. 1, pp. 31–72, 2008.
[11] P. Austin, "An introduction to propensity score methods for reducing the effects of confounding in observational studies," *Multivariate behavioral research*, vol. 46, no. 3, pp. 399–424, 2011.
[12] E. A. Stuart, "Matching methods for causal inference: A review and a look forward," *Statistical science: a review journal of the Institute of Mathematical Statistics*, vol. 25, no. 1, p. 1, 2010.
[13] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of liwc2015," Tech. Rep., 2015.