SOTICS 2021

The Eleventh International Conference on Social Media Technologies, Communication, and Informatics

ISBN: 978-1-61208-899-0

October 3 -7, 2021

Barcelona, Spain

SOTICS 2021 Editors

Nitin Agarwal, University of Arkansas at Little Rock, USA

Oana Dini, IARIA, USA/EU

SOTICS 2021

Forward

The Eleventh International Conference on Social Media Technologies, Communication, and Informatics (SOTICS 2021), held on on October 3 - 7, 2021 in Barcelona, Spain, was an event on social eco-informatics, bridging different social and informatics concepts by considering digital domains, social metrics, social applications, services, and challenges.

The systems comprising human and information features form a complex mix of social sciences and informatics concepts embraced by the so-called social eco-systems. These are interdisciplinary approaches on social phenomena supported by advanced informatics solutions. It is quit intriguing that the impact on society is little studied despite a few experiments. Recently, also Google was labeled as a company that does not contribute to brain development by instantly showing the response for a query. This is in contrast to the fact that it has been proven that not showing the definitive answer directly facilitates a learning process better. Also, studies show that e-book reading takes more times than reading a printed one. Digital libraries and deep web offer a vast spectrum of information. Large scale digital library and access-free digital libraries, as well as social networks and tools constitute challenges in terms of accessibility, trust, privacy, and user satisfaction. The current questions concern the trade-off, where our actions must focus, and how to increase the accessibility to eSocial resources.

We take here the opportunity to warmly thank all the members of the SOTICS 2021 technical program committee, as well as all of the reviewers. We also kindly thank all the authors who dedicated much of their time and effort to contribute to SOTICS 2021.

We also gratefully thank the members of the SOTICS 2021 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that SOTICS 2021 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of social eco-informatics.

SOTICS 2021 Steering Committee

Marie Sjölinder, RISE - Research Institutes of Sweden, Sweden
Elina Michopoulou, University of Derby, UK

SOTICS 2021 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Lorena Parra, Universitat Politecnica de Valencia, Spain

SOTICS 2021

Committee

SOTICS 2021 Steering Committee

Marie Sjölinder, RISE - Research Institutes of Sweden, Sweden
Elina Michopoulou, University of Derby, UK

SOTICS 2021 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Lorena Parra, Universitat Politecnica de Valencia, Spain

SOTICS 2021 Technical Program Committee

Safa'a AbuJarour, University of Potsdam, Germany
Millicent Akotam Agangiba, University of Mines and Technology, Tarkwa, Ghana
William Akotam Agangiba, University of Mines and Technology, Tarkwa, Ghana
Federico Martín Alconada Verzini, Universidad Nacional de La Plata, Argentina
Samer Al-Khateeb, Creighton University, US
Hassan Atifi, Troyes University of Technology (*UTT*), France
Zine-Eddine Baarir, Biskra University, Algeria
Qanita Bani Baker, Jordan University of Science and Technology, Jordan
Grigorios N. Beligiannis, University of Patras, Greece
Hernane Borges de Barros Pereira,Centro Universitário Senai Cimatec, Brazil
Christos Bouras, University of Patras, Greece
James Braman, Community College of Baltimore County, USA
Miguel Carvalho, INESC-ID Lisboa, Portugal / Coordinating Center for Communications and Information and Innovation Technologies, Regional Government of the Azores
K C Chan, University of Southern Queensland, Australia
Dickson K.W. Chiu, University of Hong Kong, Hong Kong
Joshua Chukwuere, North-West University (NWU), South Africa
Taiane Ritta Coelho, Federal University of Parana, Brazil
Dimitri Demergis, Rowan University, USA
Vasily Desnitsky, SPIIRAS, Russia
Nicolás Díaz Ferreyra, University of Duisburg-Essen, Germany
MeiXing Dong, University of Michigan, USA
Arianna D'Ulizia,National research council of Italy - IRPPS Research, Italy
Ritam Dutta, Surendra Institute of Engineering & Management | Maulana Abul Kalam Azad University of Technology, West Bengal, India
Abderrahim El Amine, University of Technology of Troyes, France
Luis Enrique Sánchez Crespo, Universidad de Castilla-La Mancha, Spain
Larbi Esmahi, Athabasca University, Canada
Karina Frick, University of Zurich, Switzerland

Cristian Rusu, Pontificia Universidad Católica de Valparaíso, Chile
Ali Shahrabi, Glasgow Caledonian University, Scotland, UK
Vivek Shandilya, Jacksonville University,USA
Marianna Sigala, University of South Australia, Australia
Anurag Singh, National Institute of Technology, Delhi, India
Marie Sjölinder, RISE - Research Institutes of Sweden, Sweden
Evaggelos Spyrou, University of Thessaly | NCSR Demokritos, Greece
Raquel Trillo-Lado, University of Zaragoza, Spain
Lorna Uden, Staffordshire University, UK
Stefanos Vrochidis, ITI-CERTH, Greece
Gang Wang, Hefei University of Technology, China
Huadong Xia, Microstrategy Corporation, USA
Chenwei Zhang, The University of Hong Kong, Hong Kong

Table of Contents

Community Interaction Optimization on Twitter for People with Mood Disorders

Yuichi Okada
Graduate School of Informatics
Kansai University
Osaka, Japan
email: k393424@kansai-u.ac.jp

Naoya Ito
Graduate School of Informatics
Kansai University
Osaka, Japan

Tomoko Yonezawa
Faculty of Informatics
Kansai University
Osaka, Japan
email: yone@kansai-u.ac.jp

***Abstract*—This paper proposes our designed system to estimate the optimal social networking connections for people with mood disorders, such as depression. We collected data from Twitter and analyzed users' characteristics by adopting an emotional polarity value index. Based on these data analyses, we defined each user's positivity level and estimated the level of mood disorder from the content of their tweets. We also simulated the system's use by people with severe mood disorders. A computational model based on a knapsack problem, a combinatorial problem to solve using an optimization method, was created based on the hypothesis that people likely to have mood disorders will more likely connect with people who have similarly severe mood disorders. The proposed system solved it using an approximate solution method that can be computed in a practical amount of time. As a result, (1) the user preferred a user with the same mood disorder severity when a user connected to a user with frequent tweets and (2) the difference in mood disorder severity was not as important when the user connected to a user with infrequent tweets.**

Keywords-Twitter; SNS; data mining; combinatorial optimization.

I. Introduction

We will present an outline of our research in this section.

A. SNS and Depression

Social Networking Services (SNSs) have attracted attention as a communication platform less affected by time and space constraints compared to face-to-face communication. However, the cost of SNSs' convenience is that many people suffer from physical and mental fatigue. Fatigue caused by information overload has been studied in large social networks [1].

Depression can be caused by diverse factors, including problems with health, family, economics, and employment and labor, but a relationship between social network use and depression has also been recently identified [2]. In other words, information overload emotional overload must both be addressed.

Depressed people tend to underestimate their social skills and avoid communicating with others [3]. A relationship between loneliness and depression has been suggested, and appropriate communication may help alleviate depressive symptoms [4]. Bessiere et al. [5] also pointed out the importance of considering the psychological impact of communicating on the Internet.

B. Motivation

One negative effect of social networking sites is people's inevitable unpleasant experiences with anonymous communication. In the real world, more people suffer from depression for various reasons, and SNSs' ease of use greatly benefits society and should be utilized to solve these problems. If we can improve relationships through SNS, we can alleviate mood disorders, such as depression.

C. Optimization Problem

The number of people an individual can connect with is very small compared to the number of people who use an SNS, so it is important to know who to connect with. Finding the best connections among users is essentially a combinatorial optimization problem.

Although combinatorial optimization problems are solvable in computational theory, solving them in practical computing time is difficult when their scale becomes large. However, many solutions to restrictive problems have been studied, and the solution to the knapsack problem is practical. Optimization methods based on mathematical programming make it easier to estimate the calculation accuracy and time required compared to learning algorithms such as neural networks.

Therefore, these optimization methods are suitable for building elaborate computational models. In this study, we use the surrogate constraint method for the multi-constraint nonlinear knapsack problem to optimize connections among Twitter users.

D. Research Subject

1) Language: Considering the author's language-dependent environment and the possibility of future experiments on subjects, we targeted SNS users who use Japanese. Notably, Japanese has no explicit word separators, so a morphological analysis is needed to generate word list data from sentence data for analyses.

2) Types of SNS: Several types of SNSs exist, and we chose Twitter for its wide usage rate among Japanese people and its easy-to-use an Application Programming Interface(API). The most popular SNS in Japan is LINE, but LINE is mainly used for private purposes, so it was excluded from the study.

E. Proposed System

We devised a system that suggests optimal connections for Japanese Twitter users likely to have mood disorders. Based

 ISBN: 978-1-61208-899-0

on a user's tweet data, we defined the user's characteristics, created a computational model of the knapsack problem by clustering the characteristics, and calculated a sub-optimal solution using an approximate solution method.

In general, mood disorders refer to conditions such as depression and bipolar disorder, which are characterized by emotional ups and downs so severe or unstable that they interfere with daily life. This study concerned situations ranging from weak to severe tendencies but did not use a clinical diagnosis.

Instead, we defined "mood disorder" broadly to cover mild conditions, such as lethargy and depression to severe conditions with a high possibility of depression. These states were estimated via statements posted on SNS.

F. Structure of The Paper

The paper is organized as follows: Section2 gives related works corresponding to our study. In Section 3, we show the overview of proposed system. In Section 4, the features of the Twitter users targeted for research are explained. Section 5 describes how to quantify the features of Twitter users and how to evaluate them. In Section 6, we present our method for finding the optimal interactions of Twitter users based on evaluation method we defined. In Section 7, we show the results of computer simulations using the method set up in Section 6. In Section 8, we provide a discussion of the simulation results, and in Section 9, we conclude and discuss future work.

II. Related Works

In this section, we will present the effects on the mind and the Twitter-based recommendation system as they relate to our research.

A. Effects on the mind

Tsukawa et al. [6] studied data from remarks on Twitter to diagnose depression. They used the Zung score [7], a depression evaluation index obtained using an advance questionnaire, and performed a multiple-regression analysis using the frequency of words detected by a morphological analysis of tweets from 50 subjects. Among detected words, negative words showed a high correlation with the Zung score. A moderate correlation (correlation coefficient $R = 0.45$) appeared between the score obtained by the regression equation and the Zung score. Considering the number of samples, this result seems valid, so it should be possible to diagnose depression symptoms using tweets.

One study of college students indicated that a long usage history or a high degree of dependence made people seem unhappier than others [8]. Researchers also stated this tendency would be strong among unacquainted people. On the other hand, almost no relationship exists between SNS usage duration and stress or depression when targeting SNSs while excluding Instagram [9]. The stress load depends on the quality of communication with other people.

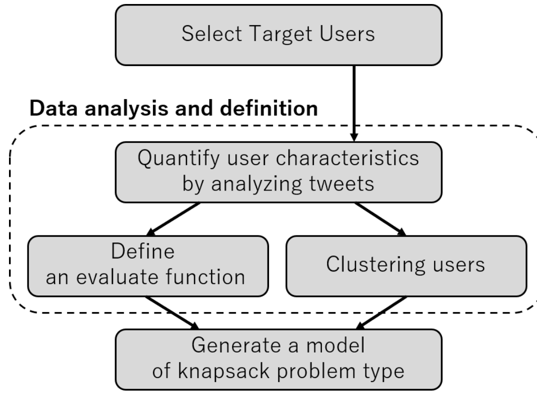

Fig. 1. System flow.

B. Twitter-based recommendation systems

Chen et al. [10] proposed a system that recommends useful tweets to Twitter users. The system provides collaborative-filtering-based recommendations via tweet content, social relationships, and other explicitly defined user data generally used as indicators of a tweet's usefulness. This system provided better recommendations than previous methods.

Research focused on followers has also indicated that follower-adding patterns can help predict who the user will follow next [11]. Another system considered users' emotions in recommendations focused on users' interests [12]. The researchers implemented better recommendations for the user by including functions with positive, negative, and neutral parameters in the evaluation axis.

All these studies aimed to efficiently select useful information for users from the large amount of information flooding the SNS.

III. System overview

In the proposed system, we obtain information from target users whose tweets contain keywords related to depression. The tweets are then analyzed and mood disorder levels are assigned to individual users in three phases. The first phase is conducting a morphological analysis of the tweets to determine the emotional polarity value. The second is quantifying the user's characteristics based on the emotional polarity value. The third is clustering users to improve the optimization by grouping users with similar characteristics. The second and third phases depend on the first phase's results, but they are not mutually dependent. Finally, these data are formulated as a knapsack problem, and the optimal solution is computed. Figure 1 shows the flow of our method.

IV. Target users

Target users include the following words in their tweets:

1) "languid" or "melancholic"
2) medicine names related to depression

The number of users who included the keywords "languid" ('da-ru-i' in Japanese) or "melancholic" ('yu-u-tsu' in Japanese) in their tweets was insufficient, so the names of 27 drugs related to mood disorder treatment, such as antidepressants, mood stabilizers, and anxiolytics, were added to the keywords to collect data. Twitter API was used for data collection, and a maximum of 3200 tweets per user (upper limit) were obtained. For each level of mood disorder, 500 user accounts were collected. Data were collected in February 2019.

V. Data analysis and definition

In this section, we will present a method for quantifying and evaluating the features of Twitter users.

A. Emotional polarity of tweets

The tweets were split into word data for each noun of a sentence using MeCab [13], a morphological analyzer. Japanese tweets must be morphologically analyzed because Japanese sentences have no explicit separators, such as spaces in English. Word data are quantified using the emotion polarity table [14], a database generated by a system that estimates a word's emotional polarity (desirability or non-desirability), with a floating number from -1 to +1 associated with the word. Values close to +1 are generally more desirable, and values close to -1 are less desirable; values close to 0 are neutral. Words not included in this database are considered 0.

B. Evaluation function of user's emotional polarity

1) Classification of mood disorder level: Based on the emotional polarity values, a user's mood disorder level was set from 0 to 3, and the user was classified into one of four classes. Research on mood disorders, especially depression, is active and many diagnostic scales exist for them. However, most scales are subjective, use questionnaires, and are not appropriate for objective assessments based on tweet data.

Non-anonymous computer mediated communication (CMC) has less effect on users than does face-to-face communication because CMC has fewer channels compared to communication in real spaces [15]. However, people with mood disorders are influenced by others' remarks, especially negative remarks. Thus, positivity should be transmitted differently depending on users' symptom levels. We therefore classified the mood disorder states according to the rules in Figure 1.

The classification involved the following steps. First, target users were divided into two groups. Group A's tweets contained names of medications for mood disorders, and Group B's tweets did not. Then, each group was further divided into two groups. In Group A, users whose tweets contained two or more names of medications for mood disorders were classified as Level 3, the highest level. The remaining Group A users were classified as Level 2, the second highest level of mood disorder. Among the members of Group B, users whose tweets contained the words "languid" and "melancholic" less than 3% of the total tweets were classified as Level 0, the lowest level of mood disorder. The remaining Group B users were classified as Level 1, the second lowest level. Figure 2 shows the flow of users.

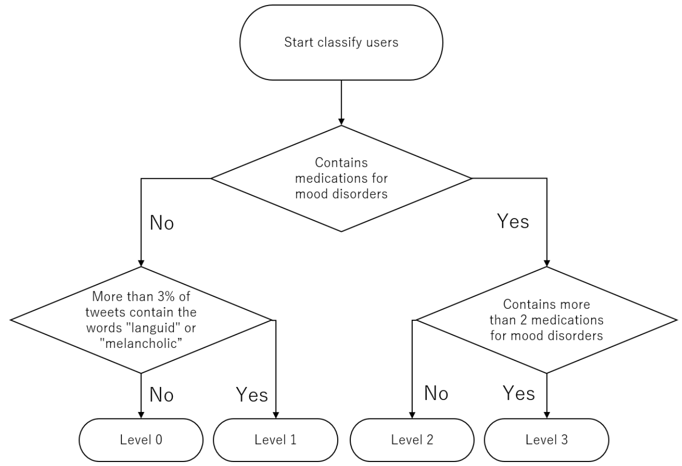

Fig. 2. The classification of users.

2) Defining the evaluation function: We defined the following evaluation function based on the hypothesis that a user with a tendency toward a mood disorder would be suitable for communicating with a user who has a similar mood disorder level and who offers many positive comments.

$$Pos(k,m) = \frac{pos_k}{neg_k}\{1.0 + \alpha * (n_k - m)\}$$

where

k : Index of user
pos_k : Positive value of $user_k$
neg_k : Negative value of $user_k$
n_k : Mood disorder level of $user_k$
m : Mood disorder level of main user

This evaluation function is named the positive function, which quantifies the degree of individual positivity and considers differences in mood disorder levels. α is a constant and is specified, so the function's value becomes positive. The values of the function $1.0 + \alpha(n - m)$ for $\alpha = 1/4$ and for $1/6$ appear in Figure 3.

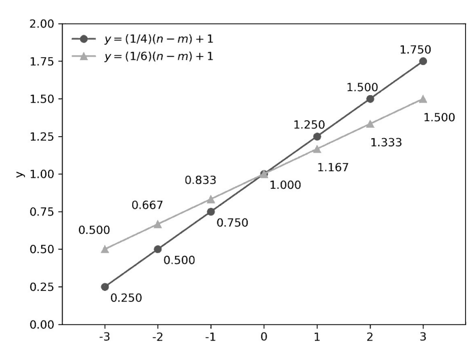

Fig. 3. Effect of level difference on positive values.

VI. Optimization

In this section, we will present a method for optimizing Twitter users' interactions using the knapsack problem.

A. Combinatorial Optimization Problems

Combinatorial optimization cannot be solved in a practical amount of computation time for a large-scale problem. This is especially critical for huge social networking sites like Twitter. For the knapsack problem among combinatorial optimization problems, however, a method to quickly obtain an approximate solution using the surrogate constraint method has been developed [16]. The surrogate constraint method generates a single constraint function by weighting multiple constraint functions, and a method for calculating the optimal weights has also been established [17]. We used this calculation method to quickly obtain an approximate solution to the single constrained knapsack problem via the algorithm [18].

B. Nonlinear knapsack problem

The multi-constraint nonlinear knapsack problem is formulated as follows.

$$\text{maximize} \sum_{i \in N} f_i(x_i)$$

subject to

$$g_j(x) = \sum_{i \in N} g_{ji}(x_i) \leq \ b_j (j = 1, 2, \cdots, m),$$

$$x_j \in A_j (j = 1, 2, \cdots, m),$$

where $N = \{1, 2, \ldots, n\}$ is a set of variable numbers, $x = \{x_1, x_2, \ldots, x_n\}$ and $A_i = \{1, 2, \ldots, a_i\} (i \in N)$ is the alternative item set for each variable.

C. Knapsack computation model

To calculate the optimal connections for a user, we defined a nonlinear knapsack-type computational model of Twitter user data. Each user has an emotional polarity value, and users with similar values were grouped into the same class, as shown in Figure 4.

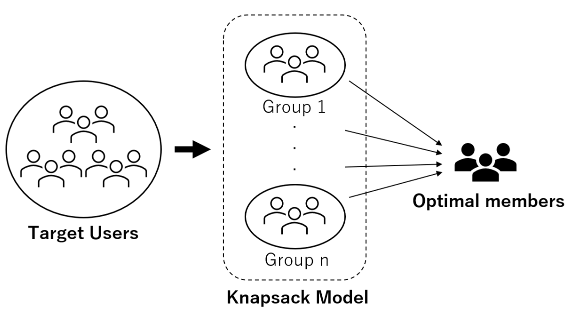

Fig. 4. Make a knapsack model.

The objective function and the inequalities for constraint function are as follows:

$$\text{maximize} \sum_{i \in N} Pos(x_i, m)$$

subject to

$$g_1(x) = \sum_{i \in N} g_1(x_i) \leq b_1$$

$$g_2(x) = \sum_{i \in N} g_2(x_i) \leq b_2$$

where

$$g_1(x_k) = \begin{cases} 1, & member\, selected\, in\, Group\, k \\ 0, & nobody\, selected \end{cases}$$

$$g_2(x_k) = days\, per\, tweet\, of\, user\, x_k$$

The function $Pos()$ is defined at V-B2. $N = \{1, 2, \ldots, n\}$ is a set of group numbers, $x = \{x_1, x_2, \ldots, x_n\}$ is a set of members of the group.b_1 is the upper limit for the number of connected users. $b2$ is the upper limit for the tweet frequency, which depends on $b1$. Thus, $b2 = (days\, per\, tweet) \times b1$.

VII. Computational Experiment

In this section, we will present a computational simulation of our proposed method.

A. Creating problem data

1) Setting the positive function: As a parameter of the positivity function, the user's mood disorder level (m) was set to 3. For α, which represents the weight of the mood disorder level on the positivity level, we set $\alpha = \frac{1}{4}$ and $= \frac{1}{6}$.

2) Setting the knapsack problem: We created the problem data according to the model defined in the previous section. A total of 2000 target users were selected, with 500 for each of the four mood disorder levels. At each level, the 500 users were sorted by the magnitude of their emotional polarity values and then divided into 50 groups of 10 users each in a descending order. It is possible to choose one person from each group or no one from any group. In other words, we created a nonlinear knapsack problem with 50 (=500/10) variables and with 11 choices in each group: 10 ways to choose a member plus a way to choose no one. The whole problem comprises 50 groups with 4 levels, for a total of 200 groups.

The first constraint ($b1$) is the number of users with connections, and we created four types: 20, 40, 60, and 80. The second constraint($b2$) is the number of days per tweet(dpr)$\times b1$, where dpr is $\frac{1}{2}$ or $\frac{1}{5}$. We assume that users who tweet twice a day are casual users, and users who tweet five times a day are heavy users.

3) Total number of problems: There were 2 types of parameters for the positive function and 4($b1$)×2($b2$) combinations regarding the knapsack problem's constraints, yielding 16 total types of problems.

B. Computation results

The calculations for each problem appear in tables I to table IV. * 1: not mood disorder, * 2: early mood disorder, * 3: mild mood disorder, and * 4 severe mood disorder.

 ISBN: 978-1-61208-899-0

TABLE I
CONNECTING HEAVY USERS($m = 3\ \alpha = 1/4$).

20	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	1	2	16	19
40	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	2	4	33	39
60	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	2	8	50	60
80	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	2	27	51	80

TABLE II
CONNECTING HEAVY USERS($m = 3\ \alpha = 1/6$).

20	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	2	2	16	20
40	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	2	8	30	40
60	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	3	17	40	60
80	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	0	3	30	47	80

TABLE III
CONNECTING CASUAL USERS($m = 3\ \alpha = 1/4$).

20	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	1	1	5	13	20
40	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	2	6	16	16	40
60	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	1	20	19	20	60
80	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	2	24	22	23	71

TABLE IV
CONNECTING CASUAL USERS($m = 3\ \alpha = 1/6$).

20	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	1	2	7	10	20
40	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	2	8	15	15	40
60	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	5	18	19	18	60
80	level0*[1]	level1*[2]	level2*[3]	level3*[4]	total
	5	22	22	21	70

VIII. DISCUSSION

In this section, we will discuss the results of computational simulations.

A. Computational results

We proposed a method for optimizing a user's connections with other users according to the mood disorder level. Based on the optimization results, the optimal solution is 1) to connect users with a similar mood disorder level when a user desires connections with heavy users, and 2) to connect with others at various levels from 1 (early mood disorder) to 3 (severe mood disorder) when the user desires connections with casual users.

In connecting with heavy users, we anticipate that the amount a user communicates with a particular person will increase and that the user will spend much time browsing Twitter to see what they are up to. In other words, if users want to communicate deeply for a long time, they should engage with people at a similar mood disorder level. In the case of $\alpha = \frac{1}{6}$, the above characteristics appeared in users preferring to connect with casual users and with heavy users. In the case of a larger $\alpha (= \frac{1}{4})$, however, these characteristics appeared stronger.

A comparison between heavy users and casual users shows that users with severe mood disorders (Level 3) often tweet frequently, and many Level-0 users had fewer tweets per day than Level-3 users. On the other hand, users who connect with casual users have shorter visits to Twitter. Therefore, connecting with people of various mood disorder levels is considered optimal. This observation aligns with the difference between using Twitter for communication and using it for information gathering.

When we want to get to know other people deeply, we gain a sense of security by connecting with people in situations similar to ours. When we want a shallow relationship with people, we gain diverse ideas by connecting with people who have diverse situations and personalities.

B. About the system

1) Knapsack problematization: Although combinatorial optimization problems are generally hard to compute, we quickly obtained a solution using an approximate solution method in the knapsack problem. This should be sufficient for developing applications that run at a practical speed. In this study, we used the positive function value to group the users. Various clustering methods can be used to group users via indicators such as personality and knowledge traits, which will increase the scope of the analysis.

2) Evaluation function: In this study, we defined a positive function as an evaluation function. The difference in the computational results between heavy users and casual users suggests the positive function worked well.

However, the sentiment polarity dictionary required for the positive function was created based on WordNet [19] words, so the performance of the dictionary may not be sufficient in the SNS world, where abbreviations and new words are created every day (such as on Twitter).

Developing a sentiment polarity dictionary based on word variance information on Twitter should be considered to improve evaluation function. Because of the wide variety of user characteristics and purposes of use in practice, multiple evaluation axes [20] should be considered instead of using a single polarity of positive/negative.

C. Limitation

We used the Twitter API's keyword search to obtain users' data. The first search did not provide sufficient information, so search words were appended later. In this regard, the data collection method is somewhat inconsistent.

We conducted an additional study to estimate users with depressive tendencies and obtained data from 1,008,618 profiles from the followers and friends of 100 tweeters. The data were obtained from a keyword search for "depression" ('u-tsu' in Japanese Kanji). The number of users who had sentences in their profiles that could indicate they were depressed was 882.

By more carefully observing their tweets and their interaction relationships, we should make more accurate estimates of mood disorder levels. Correlations between mood disorder levels and positive functions, as well as other statistical analyses, must be studied to show a valid estimation method.

D. Summary

Using this study's proposed system, we show the possibility of designing an optimal way for users to connect with other users. This system reveals to users the best way to interact with other people.

If this approach can be realized, the system could reduce mood disorders and provide support for healthy reintegration into society by using the online resource of SNS.

IX. Conclusion And Future Work

In this paper, we analyzed tweet data from about 2,000 Twitter accounts and calculated positive and negative values for individual users using an emotional polarity dictionary. We also estimated the mood disorder degree using the vocabulary in the tweets and classified the users into four mood disorder levels. A positive function was also defined based on the assumption that statements from other users with similar or more severe mood disorders would positively affect the user.

Next, we created a computational model of the knapsack-problem type using these characteristics. We then simulated the optimal community environment for users with mood disorders by solving the problem. From the results, we assumed that a user with a high-level mood disorder who wanted to interact with users who tweeted frequently should optimally follow people with the same level of mood disorder as the user. On the other hand, we inferred that a user with a high-level mood disorder who wanted to interact with casual users who tweeted infrequently should optimally connect with people who have various levels of mood disorder.

These estimations can be made in a practical computation time, suggesting the possibility of developing an interactive user-recommendation system. Considering actual use, however, a unified evaluation function is inappropriate because individual users have different purposes for using SNS. Therefore, a system that dynamically generates an evaluation function according to a user's purpose and values is necessary.

For future work, we want to examine our proposed system's validity and consider systems that can help people with mood disorders. For users without mood disorders, we should also research SNS uses that allow them to enjoy their social networking life to the fullest while maintaining their mental health. Slandering behavior, flaming, and polarization are major social issues in modern SNS use. SNSs are a powerful tool for knowledge acquisition and communication, and it is important to use SNSs well instead of running from them.

Acknowledgment

This research is supported in part by JSPS KAKENHI 19H04154, 21K11968, and 19K12090.

References

[1] A. Dhir, P. Kaur, S. Chen, and S. Pallesen, "Antecedents and consequences of social media fatigue," *International Journal of Information Management*, vol. 48, pp. 193–202, 2019.

[2] S. Cunningham, C. C. Hudson, and K. Harkness, "Social media and depression symptoms: A meta-analysis," *Research on Child and Adolescent Psychopathology*, vol. 49, no. 2, pp. 241–253, 2021.

[3] C. Segrin, "Social skills deficits associated with depression," *Clinical psychology review*, vol. 20, no. 3, pp. 379–403, 2000.

[4] A. R. Rich and M. Scovel, "Causes of depression in college students: A cross-lagged panel correlational analysis," *Psychological Reports*, vol. 60, no. 1, pp. 27–30, 1987.

[5] K. Bessière, S. Kiesler, R. Kraut, and B. S. Boneva, "Effects of internet use and social resources on changes in depression," *Information, Community & Society*, vol. 11, no. 1, pp. 47–70, 2008.

[6] S. Tsugawa, Y. Mogi, Y. Kikuchi, F. Kishino, K. Fujita, Y. Itoh, and H. Ohsaki, "On estimating depressive tendencies of twitter users utilizing their tweet data," 03 2013, pp. 1–4.

[7] W. W. Zung, "A self-rating depression scale," *Archives of general psychiatry*, vol. 12, no. 1, pp. 63–70, 1965.

[8] H.-T. G. Chou and N. Edge, "'they are happier and having better lives than i am': The impact of using facebook on perceptions of others' lives." *CyberPsychology, Behavior Social Networking*, vol. 15, no. 2, pp. 117 – 121, 2012.

[9] E. Donnelly and D. Kuss, "Depression among users of social networking sites (snss): The role of sns addiction and increased usage," *Journal of Addiction and Preventive Medicine*, vol. 1, no. 2, p. 107, 2016.

[10] Chen *et al.*, "Collaborative personalized tweet recommendation," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 661–670.

[11] M. Islam, C. Ding, and C. Chi, "Personalized recommender system on whom to follow in twitter," in *2014 IEEE Fourth International Conference on Big Data and Cloud Computing*, Dec 2014, pp. 326–333.

[12] G. Sansonetti, D. F. Gurini, F. Gasparetti, and A. Micarelli, "Dynamic social recommendation," in *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2017, pp. 943–947.

[13] T. Kudo, "Mecab: Yet another part-of-speech and morphological analyzer," *https://sourceforge.net/projects/mecab/*, 2006, retrieved: 08, 2021.

[14] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orientations of words using spin model," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ser. ACL '05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 133–140. [Online]. Available: https://doi.org/10.3115/1219840.1219857

[15] K. Sassenberg, M. Boos, and S. Rabung, "Attitude change in face-to-face and computer-mediated communication: private self-awareness as mediator and moderator," *European Journal of Social Psychology*, vol. 35, no. 3, pp. 361–374, 2005.

[16] F. Glover, "Surrogate constraints," *Operations Research*, vol. 16, no. 4, pp. 741–749, 1968.

[17] M. E. Dyer, "Calculating surrogate constraints," *Mathematical Programming*, vol. 19, no. 1, pp. 255–278, 1980.

[18] Y. O. Naoya Ito, Tomoko Yonezawa and Y. Nakagawa, "[global greedy algorithm for the nonlinear knapsack problem] hisenkei knapsack mondai ni taisuru global greedy ho (in japanese)," in *Information Processing Society of Japan Kansai Chapter Conference*, vol. 2018, sep 2018, pp. Paper ID G–05, (4 pages, in Japanese). [Online]. Available: https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_action_common_download&item_id=191790&item_no=1&attribute_id=1&file_no=1&block_id=8&page_id=13

[19] F. Bond, T. Baldwin, R. Fothergill, and K. Uchimoto, "Japanese semcor: A sense-tagged corpus of japanese," in *Proceedings of the 6th global WordNet conference (GWC 2012)*. Citeseer, 2012, pp. 56–63.

[20] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

Applicability of Social Media Elements in Notification Systems in Large Interconnected Organisations

Igor Jakovljevic
ISDS
Graz University of Technology
Graz, Austria
e-mail: igor.jakovljevic@cern.ch

Christian Gütl
ISDS
Graz University of Technology
Graz, Austria
e-mail: c.guetl@tugraz.at

Andreas Wagner
IT Department - Web Frameworks
CERN
Geneva, Switzerland
e-mail: andreas.wagner@cern.ch

***Abstract*—Access to large amounts of information has resulted in users experiencing the effects of information overload, which is the state where users are presented with a much higher amount of information than they can process at a given time. Information overload is also observed in Social Media Platforms (SMP), which has resulted in the invention of mitigation techniques for information overload. Currently, there are many different approaches to deal with information overload. One popular method to deal with information overload and minimize the time spent by users analyzing large amounts of information is using notification systems. Notifications aim to display the most relevant information at a certain point in time to the user without disturbing the daily workflow of an individual. This work focuses on determining aspects of social media that can be integrated into a notification system. These aspects have the purpose to enhance and aid information retrieval, visualization, and distribution capabilities in a notification system and reduce the effects of information overload. As part of this work, we executed a study with 35 participants and presented different use-cases of Social Media Elements (SME) to evaluate their usability. The preliminary results of the study showed that adding social media elements to notifications increased the credibility and clarity of notifications. The participants reacted positively to notifications that were formatted as social media posts, rating them more trustworthy as compared to traditional notifications. SME had the effect of aiding the participants to better determine the difference between fake and real information.**

Keywords—Social Media; Notifications; Large Organisations; Hashtags; Microblogs

I. Introduction

Adoption of modern Information and Communication Technologies (ICT) have increased the amount of generated information and made information easily accessible. However, this also has resulted in users experiencing information overload [1], which can be defined as the state when users are presented with large amounts of information, that exceed the processing capability of the users [2].

One source of information overload is notifications, since users receive a large amount of notifications from multiple applications on multiple devices (e.g., desktop notifications or mobile notifications). Notifications allow applications such as email clients, messaging applications, calendars, and others to inform users of incoming messages from other users, upcoming events, reminders, new emails, and more without explicitly requiring user interaction with the application. Since each application has a specific notification format, the user is presented with a large amount of different information, making it hard to process [3].

Based on a study of 40191 randomly selected participants from different areas of work, users receive on average 44.9 notifications per day from multiple sources. Participants received notifications from 173 applications. Some of the applications were email applications (e.g., Gmail or Outlook), text messaging applications (e.g., Whatsapp or SMS apps), and voice messaging applications (e.g., Google Hangout and Skype) [4]. Findings also shown that high number of notifications, in particular from email clients and social networking applications, correlate with increased stress and the feeling of being overwhelmed. They distract users from executing current tasks and induce negative emotions [5]. Another study based on a sample of high-performing management individuals has revealed that the increase of information overload leads to more stress and negative emotions in individuals [2]. Users acknowledge notifications as potentially disruptive and distracting since they do disrupt the current engagement of the user [4].

Despite the disruptive nature of notifications, users decide to use them because of their benefit in providing relevant information. In this context, notification systems can be beneficial and attempt to aggregate the previously mentioned information from different sources (email clients, news portals, messaging platforms, and others), and deliver it to the user in the form of notifications [6]. Besides providing information aggregation and notification delivery, notification systems enable the management of notifications (e.g., selecting which applications are allowed to send notifications), which reduces the user need to constantly interact with different applications [7]. The success of a notification system hinges on accurately supporting the user with information between tasks, while simultaneously enabling utility by providing access to additional information [8]. Notification systems attempt to keep the users informed by balancing the amount of valuable information provided and the disruption the information causes. It is necessary to find means to coordinate the delivery of notifications from multiple applications across multiple devices or/and display only relevant information at a glance. By bringing multiple sources of notifications together, the user can determine the importance of a notification and reduce the level of distraction [9].

According to [8], there are three critical parameters for the

 ISBN: 978-1-61208-899-0

creation of a successful notification system:

1) **Interruption** - is defined as an event, where the user has to shift their attention from the main task and switch focus to the notification. Examples of these events are receiving notifications while operating heavy machinery, where the notification should not distract the user from the main task. However, other situations like medical emergency alerts, require that the notification explicitly interrupts the user [10].
2) **Reaction** - is defined as the response to the stimulus provided by the notification. Some examples of user reactions are ignoring notifications, removing them from the notification list, and clicking on the notification.
3) **Comprehension** - is defined as the use of notification systems with the goal of remembering and making sense of information at a later point in time. Based on past reactions, a notification system can show notifications to the users when they are more likely to read them.

While quick and correct reaction to information is important in many situations, it is also important to present the information in a comprehensible way. Notifications should display a balance between the interruption, reaction, and comprehension parameters [8].

One of the main challenges with designing a notification system is learning when and how to display understandable and valuable messages at a glance without explicitly disturbing or distracting the user. This problem has been tackled in different disciplines. Potential practical concepts can be found in social media, especially Social Media Marketing (SMM). The goal in SMM is to present information to the user at a specific time based on previous user behavior and experiences with other similar users. The information contains social media elements and should not irritate the user but stimulate engagement with the content. This SMM information usually has the goal to guide the user to a social media site [11].

Social media sites have become one of the most popular social behaviors among humans, and recent statistics suggest that more than two-thirds of internet users use social media sites [12]. One of the main reasons for its popularity is the user engagement and personalized information it provides to the users. There are also drawbacks such as the lack of security, internet addiction, frequent interruptions from other tasks, information overload, creation of information bubbles, and loss of social contacts [13].

Social Media and SMM concepts related to user engagement and information presentation can potentially be adopted in notification systems to improve message flow to users and enhance user engagement.

Kietzmann et al. [14] identified seven main functional building blocks of social media: identity, conversations, sharing, presence, relationships, reputation, and groups. These building blocks can be identified in various social media applications, like networking sites, photo-sharing platforms, blogging platforms, video-sharing platforms, collaboration platforms, and micro-blogging platforms.

In this paper, we want to identify social media elements based on previously mentioned functional building blocks of social media and explore which of these social media elements can be adopted in a notification system. The goal is to improve user interaction and navigation, information value, information dissemination of notifications, and comprehension of notifications in notification systems. In addition, attention is also given to mitigate possible side effects of social media elements and notification systems, such as wasting time analyzing and reviewing the information provided to the user via the notification system.

Based on the observations stated above, more specifically, the main research questions are: RQ1: Which elements of social media can be integrated into notification systems to display understandable and valuable notifications at a glance without explicitly disturbing the user?, RQ2: Would users prefer to receive notifications with integrated social media elements like hashtags, topic keywords, source information, rating by other users, and groups information?, RQ3: How do users react to notifications with this additional information?, and RQ4: Which emotions do users experience when receiving notifications with and without this additional information?

To this end, the remainder of this paper is organized as follows: Section II covers the literature overview and discusses current topics in social media, notification systems, and their relation and use-cases. In Section III, the methodologies used in the study and the study are explained. Results are presented in Section IV, together with the discussion of the study outcome. Finally, we conclude the work in Section V.

II. Background and Related Work

Inspired by SMM, where the integration of social media elements into marketing information has led to increased user engagement and satisfaction, we propose adopting social media elements into notification systems and notifications [11]. In analogy to gamification applying game design elements in non-game contexts [15], it is proposed to integrate social media elements in non-social media contexts. The application in notification systems aims to improve the readability of notification and increase its informational value.

The remainder of this section assesses the drawbacks and advantages of notification systems, outlines social media elements, and investigates possible integration in notification systems.

A. Notification Systems

There are many different implementation versions of notification systems. The most commonly used are push notification systems for mobile phones, desktop status notification systems, browser-based notification systems, in-vehicle information systems, and others [16]. As mentioned in the previous chapter, notification systems attempt to convey important information to users effectively without creating an unwanted intrusion into current user tasks [6]. Selecting important information for the user is a difficult task. A study of 400+ participants has shown that users are not satisfied with the

 ISBN: 978-1-61208-899-0

notifications they receive from notification systems because they do not express the user's current interest. This leads to users ignoring most notifications from these systems [17]. Besides determining what is relevant information for the user, an essential concern in notification systems is the display of notifications without a significant interruption of users' main tasks. Visual implementations of notifications that typically are not a user's main attention priority are called secondary displays. Users willingly sacrifice brief interruptions from their primary task to view information of interest in these secondary displays [18].

B. Social Media Elements

The above-mentioned functional building blocks of social media are umbrella terms used to cover many social media elements observed on different social media platforms. Based on the analysis of social media sites and research on social media aspects [14][11][19] we identified and summarized some of the most common elements. Table I displays these elements.

TABLE I
SUMMARY OF MAIN SOCIAL MEDIA ELEMENTS

Social Media Element	**Description**
Hashtags	A hashtag is a metadata tag type used on social networks to help users find resources with a specific theme or content [19][20]
Microblogs	Microblog services allow users to post and share short textual messages that are then propagated to an audience, which can then quickly interact with the posts and between each other [21]
Content approval/ disapproval	Social cues that send signals of social appropriateness or social acceptance of content to the content creator. Examples of these social elements are Likes, Retweets, Reactions, and more [22]
User Groups	User groups represent the extent to which users can form communities and sub-communities. The more 'social' a network becomes, the bigger the group of friends, followers, and contacts.
User-to-User Relationship	User-to-user relationships express the extent to which users can relate to each other (e.g., friendships on Facebook or Followers on Twitter) [14]
User Identity	It represents the degree to which users expose their identities on social media sites. It includes exposing information such as name, age, gender, profession, location, and other users' identifiable information [14].

Considering the definition of social media elements from Table I and the description of notification systems above, we have decided to exclude user identity from our research and for the review in this section. The main reason for the exclusion of this element is that it is too focused on the individual. Including user identity information in notifications displayed to the user does not improve the information on notifications. Showing this information would be redundant for the user and could not be integrated into the context of notifications without privacy concerns.

1) Hashtags: The content of hashtags can be dynamically generated or user-generated and can only consist of letters, digits, and underscores. Thus, hashtags are iconic features that enable easy retrieval of connected resources [19][20]. They are used to construct a personal word/hashtag vector space of a user by examining the users' linguistic expression [23]. Besides identifying and representing user features, hashtags connect similar resources by assigning tags to provide contextual information [24]. According to [25], linking information on Twitter with information from other sources like Wikipedia led to increased understanding of the information and productivity when consuming the information. For example, in the context of notifications, hashtags could represent the topic of the notification or connect the notification to information related to the topic, making it easy for the user to determine if the notification is related to the current task.

2) Microblogs: Similar to microblog posts, notifications are messages displayed to the users with the intent to share information. These messages contain information from different applications (e.g., email subject and part of email text, new message alert). Based on the above description, it can be concluded that notifications share similarities to microblog post entries. However, unlike notifications which do not contain much additional information in their visual representation, microblog posts contain aspects of social media, which allow the users to determine the importance and validity of a post.Aspects like the number of individuals that have shared, liked, or approved the post, topics related to the shared post, and the type of individuals that have interacted with the post are of crucial importance to assess the value of the post and the information within [26][27].

Hashtags in Microblogging services contain information about temporal trends of the information stream and the topology of the spread of information. This makes hashtags a tool suitable for archiving, tracking, and disseminating information [28][29].

3) Content approval/disapproval: Providing and receiving feedback is a fundamental component of participation in social media. In addition, the popularity of social media has enabled the use of rich user information from Facebook and other social networks to predict users' latent traits for recommendation [30]. Based on the previously mentioned study, users have expressed a need for more personalization in notifications; integrating likes into a notification system as a means to gather feedback from the user related to the notifications could be beneficial for improving the satisfaction rate of users [17].

4) User Groups: A widely discussed relationship group metric is Dunbar's Number, proposed by Robin Dunbar in 1992. He theorized that people have a cognitive limit that restricts the number of stable social relationships with other people to about 150. Social media platforms have recognized that many communities grow well beyond this number and offer tools that enable users management of memberships [31]. The assumption that the vocabulary used to discuss a topic stays similar between different user communities and does not vary significantly over time directly suggests that it is possible to compute the overlap of topics of two or more communities. This community similarity can connect communities from different social networks (e.g., Facebook), facilitate information sharing between communities, and extract community interest [32]. Furthermore, user groups and group behavior information infer social cues, including group information (e.g., number of people with the same interests who approved a notification or executed a specific action) in notifications could increase the credibility and information dissemination of notifications.

5) User-to-User Relationship: The type of relationships users form between each other determines what information exchanges between them. For example, when users form professional relationships online, the information exchanged between them will be of professional content and high value, compared to friendly relationships where the information is of a different nature [14]. User relationship information could be used in notification systems to determine the character of information presented to the user.

C. Discussion

Towards our goal to determine how social media elements can enrich notifications with additional information, the section above outlined vital social media elements and investigated their application for this purpose. Table II summarizes how social media elements could be beneficial for notification systems.

TABLE II
SOCIAL MEDIA ELEMENTS AND USABILITY IN NOTIFICATION SYSTEMS

Social Media Element	**Usability in Notification Systems**
Hashtags	Quick access to topic information; Enables instant classification of notifications by topic; Linking external information to the notification
Microblogs	Social Media Posts provide information representation ideas for notification due to their similarity; Content Sharing does not have a direct use in notifications
Content approval/disapproval	Provide a way for the user to express interest
User Groups	Provide additional information and credibility of information based on the opinion of a group of users
User-to-User Relationship	Provide different types of additional information based on relationships with different users

Hashtags and user group elements provide additional information, potentially enhancing the information in notifications. Integrating these elements could increase the trustworthiness of notification systems and reduce the time needed for a user to evaluate the importance of notifications. Since notification systems lack a direct user feedback mechanism, integrating content approval/disapproval elements could provide it.

For Microblogs, our research focused on two features Social Media Posts and Content Sharing. Due to the lack of applicability in notification systems, content sharing was excluded. However, considering that social media posts share similarities with notifications, we determined that formatting information in notifications similar to social media posts by including hashtags, more personalized text, and information sources, could benefit notification systems.

Even though user-to-user relationships offer great insights into users' interests, knowing the user and connections are mandatory to integrate this element into a notification system. Due to the setting of our initial study, we excluded this element from the evaluation since it was necessary to track user relationships over a more extended period.

To this end, we have selected four social media elements for evaluation based on their applicability in notification systems: hashtags, user group information, content approval/disapproval, and social media posts (formatting the content of the notification as a social media post).

III. User Study

To determine the effects of certain social media elements on individuals, we designed an online study. The user study addresses the above-defined research questions: RQ1 to RQ4.

The study was designed as an AB study including the following parts:

1) General Questionnaire: contains questions listed in Table III that aim to identify the value and effects of additional information in notifications on the user. This questionnaire aims to provide insights to RQ1 and RQ2, by explicitly asking the participants about how they perceive SMEs in the notifications they received.

TABLE III
GENERAL QUESTIONNAIRE QUESTIONS

Question
Q1: Did you find the additional information in the notification valuable?
Q2: When I received notifications with additional information I was more confident in the notification?
Q3: Rank the additional information by importance
Q4: It was easier to understand the notification when I had additional information in the notification?
Q5: Did the notification break your concentration while executing the task?

2) Article Feedback: contains questions listed in Table IV that ask participants to evaluate if the articles are fake or not. The responses determine if notifications with additional information help determine the truthfulness of articles and how users react to notifications with additional information.

TABLE IV
ARTICLE FEEDBACK QUESTIONS

Question
Q1: Do you think that the article "Friends Reunion" is Fake or Real?
Q2: Do you think that the article "Instagram for Children" is Fake or Real?
Q3: Do you think that the article "People live in a 3D-Printed House" is Fake or Real?
Q4: Do you think that the article "3 Reasons Why You Should Stop Eating Peanut Butter Cups!" is Fake or Real?
Q5: Do you think that the article "Us Bacon Reserves Hit 50 Year Low" is Fake or Real?

3) Computer Emotion Scale (CES): used to answer RQ4 by determining the emotional influence of notifications on the participants since it provides one of the most scientific ways for emotion evaluation [33].

4) System Usability Scale (SUS): used to determine if the participants would prefer to receive notifications with additional information, which directly correlates with RQ2 and RQ3. It provides a trustworthy evaluation tool for usability testing [34].

The study was created using the CoDiS Survey Tool [35], which tracked and analyzed participants' behavior and presented specific assignments. The CoDiS Survey Tool is a web based evaluation tool. The participants were asked to read articles, mentioned in Table V and execute predefined tasks (share articles, comment on the article, and more). As the participants were doing these tasks, notifications related to the articles were displayed. Notifications were displayed as part of the CoDiS Survey Tool as web elements that appear when the user starts reading an article. Depending on the user group, these notifications were either with additional information or without additional information. The additional information included hashtags, user group information, and social media post formatting. This additional information integrates all selected social media elements from the previous chapter.

TABLE V
ARTICLE TITLE AND VALIDITY

#	Title	Is Fake
1	Friends Reunion	No
2	People live in a 3D-Printed House	No
3	Instagram for Children	No
4	US Bacon Reserves Hit 50 Year Low	Yes
5	3 Reasons Why You Should Stop Eating Peanut Butter Cups!	Yes

The participant target groups for the study were high school and university students. In total, 215 individuals were asked to participate, and only 35 completed the study. The age of the participants varied from 15 to 34 years old, with 57.14% of the participants being in the range from 15 to 20 years, 25.72 % being in the range 20-25, 14.289 % in the range 25-30, and 2.85% in the range above 30 years old. Female participants made 28.57% of the total amount of participants, while male participants made 71.43%. As stated previously the study was designed as an AB study, this is why the participants were divided into two groups (Group A and Group B). The purpose of this division is to reduce bias between users. Both groups received the first article with additional information notifications. The purpose of this was to create a control article and familiarize the users with this type of notifications. Group A received simple notifications on even-numbered articles, while group B received them on odd-numbered articles. After the participants finished reading the articles and the article-related tasks, they had to complete an evaluation.

IV. Findings and Discussion

Analyzing the answers to questions presented in Table III, we have concluded that the participants find notifications easier to understand and share the thought that they have more credibility when presented with additional information. The additional information in notifications has increased the value of notifications to the user based on answers to Q1 from Table III, where 85.71% confirmed the premise. The participants had more confidence in notifications with additional information in comparison to formal notifications, based on answers to Q2 from Table III. Based on Q3 from Table III 77.14% of the participants stated that they find it easier to understand notifications with additional information. With 60% of participants answering with ”Yes” to Q5 from Table III, we can confirm that notifications break user concentration, which validates results of previous research [4][2].

According to [8], the success of notification systems is dependent on the information they convey to the user. The survey participants agree with this as shown in Table VI. It reveals that users care predominantly about the content and source of notifications. It implies that adding additional information to validate the content and source increases their value to users. The results in Table VI also validate our proposal that formatting notifications as social media posts could improve the information presented to the user since the content was formatted to be similar to a social media post. Contrary to our research, group information (e.g., ”22 readers validated text”) was not ranked as highly important by the participants.

TABLE VI
ADDITIONAL INFORMATION RANKING BY IMPORTANCE

Additional Information	Very Important	Not at all important
Information Source	10 (33.33%)	1 (3.33%)
Hashtags	2 (6.67%)	4 (13.33%)
Content of the Notification	10 (33.33%)	1 (3.33%)
Group or Reader Validation Info (e.g., ”22 Readers Validated Text”)	1 (3.33%)	10 (33.33%)
Notification Position	7 (23.33%)	14 (46.67%)

The distribution of SUS answers reveals that most of the users agree or strongly agree with questions Q1, Q3, Q4, Q5, and Q7 of the SUS [34] while disagreeing or strongly disagreeing with the rest of the questions. Due to a large number of neutral answers, the average rating of the scale is 69.78. This is slightly above the limit of 68 set by [34] as the value that is the minimum for a usable system. Based on the results of the SUS, we can infer that the users would prefer to use a notification system with social media elements.

TABLE VII
PERCENTAGE AN ANSWERS HAS BEEN SELECTED ON THE COMPUTER EMOTION SCALE

	None of the Time	Some of the Time	Most of the Time	All of the Time
Happiness	20.95%	31.43%	24.76%	22.86%
Sadness	68.57%	24.29%	4.29%	2.86%
Anxiety	69.29%	18.57%	8.57%	3.57%
Anger	65.71%	21.90%	7.62%	4.76%

The result of the CES is shown in VII, the table contains a list of feelings a participant has experienced. The CES shows that the users were happy most of the time executing tasks and receiving notifications, while none of the time experiencing sadness, anxiety, and anger. Based on VII the emotion anxiety has the lowest score because most of the users rated it with ”none of the time” followed by Sadness and Anger. The best-rated emotion was Happiness where the majority of the users answered with either ”Some of the Time”, ”Most of the Time” or ”All of the Time”. These results do not correlate with previous studies, where users experienced negative emotions and stress while receiving notifications [5].

The participants were asked to determine if the articles

 ISBN: 978-1-61208-899-0

they read were fake or not. They rated correctly in 57.93% of the cases. The participants that were shown notifications with additional information, selected fake and real news with a 6.61% greater accuracy.

Due to the inability to track the usage of notifications over a more extended period, we could not evaluate all social media elements.

V. Conclusion

In conclusion, this research study displays the potential of social media elements in different disciplines, focusing on uses in notification systems. The initial research study shows how the selected social media elements can potentially increase user satisfaction and the value of information in notification systems. Since time constraints were an issue and restricted the number of SMEs that could be evaluated, future work might include an analysis of user reactions to notifications with additional information over a prolonged period. This would enable a better evaluation of the analyzed SME and additional SMEs that could not be part of this study. Prolonged tracking of user reactions to different combinations of SMEs in notifications might lead to a novel approach in the use of SMEs and notification systems. The survey results could provide the initial steps towards new use cases of social media applications in notification systems and other disciplines.

References

[1] C. Gunaratne, N. Baral, W. Rand, I. Garibay, C. Jayalath, and C. Senevirathna, "The effects of information overload on online conversation dynamics," *Computational and Mathematical Organization Theory*, vol. 26, pp. 1–22, Jun. 2020. doi: 10.1007/s10588-020-09314-9.

[2] P. G. Roetzel, "Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework developmen," *Business Research*, vol. 12, no. 2, pp. 479–522, Dec. 2019. doi: 10.1007/s40685-018-0069-z.

[3] A. Visuri, N. van Berkel, T. Okoshi, J. Goncalves, and V. Kostakos, "Understanding smartphone notifications' user interactions and content importance," *International Journal of Human-Computer Studies*, vol. 128, pp. 72–85, 2019, issn: 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2019.03.001.

[4] A. S. Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale assessment of mobile notifications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14, Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 3055–3064, isbn: 9781450324731. doi: 10.1145/2556288.2557189.

[5] S. T. Iqbal and E. Horvitz, "Notifications and awareness: A field study of alert usage and preferences," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '10, Savannah, Georgia, USA: Association for Computing Machinery, 2010, pp. 27–30, isbn: 9781605587950. doi: 10 . 1145 / 1718918 . 1718926.

[6] D. S. McCrickard, M. Czerwinski, and L. Bartram, "Introduction: Design and evaluation of notification user interfaces," *International Journal of Human-Computer Studies*, pp. 509–514, 2003, issn: 1071-5819. doi: https://doi.org/10.1016/S1071-5819(03)00025-9.

[7] M. Pielot, K. Church, and R. de Oliveira, "An in-situ study of mobile phone notifications," in *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices Services*, ser. MobileHCI '14, Toronto, ON, Canada: Association for Computing Machinery, 2014, pp. 233–242, isbn: 9781450330046. doi: 10.1145/2628363.2628364.

[8] D. McCrickard, C. Chewar, J. Somervell, and A. Ndiwalana, "A model for notification systems evaluation—assessing user goals for multitasking activity," *ACM Trans. Comput.-Hum. Interact.*, vol. 10, pp. 312–338, Dec. 2003. doi: 10.1145/966930.966933.

[9] D. Weber, A. S. Shirazi, and N. Henze, "Towards smart notifications using research in the large," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, ser. MobileHCI '15, Copenhagen, Denmark: Association for Computing Machinery, 2015, pp. 1117–1122, isbn: 9781450336536. doi: 10.1145/2786567.2794334.

[10] D. C. Mcfarlane and J. L. Sibert, "Interruption of people in human-computer interaction," Ph.D. dissertation, 1998, isbn: 0599230495.

[11] F. F. Li, J. Larimo, and L. Leonidou, "Social media marketing strategy: Definition, conceptualization, taxonomy, validation, and future agenda," *Journal of the Academy of Marketing Science*, vol. 49, pp. 51–70, Jun. 2020. doi: 10.1007/s11747-020-00733-3.

[12] K. Kircaburun, S. Alhabash, Ş. Tosuntaş, and M. Griffiths, "Uses and gratifications of problematic social media use among university students: A simultaneous examination of the big five of personality traits, social media platforms, and social media use motives," *International Journal of Mental Health and Addiction*, vol. 18, no. 3, pp. 525–547, 2020.

[13] M. Drahošová and P. Balco, "The analysis of advantages and disadvantages of use of social media in european union," *Procedia Computer Science*, vol. 109, pp. 1005–1009, 2017, 8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 16-19 May 2017, Madeira, Portugal, issn: 1877-0509. doi: https://doi.org/10.1016/j.procs.2017.05.446.

[14] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? get serious! understanding the functional building blocks of social media," *Business Horizons*, vol. 54, no. 3, pp. 241–251, 2011, issn: 0007-6813. doi: https://doi.org/10.1016/j.bushor.2011.01.005.

[15] D. Dicheva, C. Dichev, G. Agre, and G. Angelova, "Gamification in education: A systematic mapping study," *Educational Technology Society*, vol. 18, pp. 75–88, Jul. 2015.

[16] D. McCrickard and C. Chewar, "Designing attention-centric notification systems: Five hci challenges," in *Cognitive Systems: Human Cognitive Models in Systems Design*, Psychology Press, 2006, ch. III, pp. 67–89.

[17] S. Pradhan, L. Qiu, A. Parate, and K. Kim, "Understanding and managing notifications," in *"IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9. doi: 10.1109/INFOCOM.2017.8057231.

[18] D. S. McCrickard, R. Catrambone, C. M. Chewar, and J. T. Stasko, "Establishing tradeoffs that leverage attention for utility: Empirically evaluating information display in notification systems," *International Journal of Human-Computer Studies*, vol. 58, no. 5, pp. 547–582, 2003, issn: 1071-5819. doi: https://doi.org/10.1016/S1071-5819(03)00022-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581903000223.

[19] D. Correa and A. Sureka, "Mining tweets for tag recommendation on social media," in *Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents*, ser. SMUC '11, Glasgow, Scotland, UK: Association for Computing Machinery, 2011, pp. 69–76, isbn: 9781450309493. doi: 10.1145/2065023.2065040.

[20] J. Bieniasz and K. Szczypiorski, "Methods for information hiding in open social networks," *JUCS - Journal of Universal Computer Science*, vol. 25, no. 2, pp. 74–97, 2019, issn: 0948-695X. doi: 10.3217/jucs-025-02-0074.

[21] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" In *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10, Raleigh, North Carolina, USA: Association for Computing Machinery, 2010, pp. 591–600, isbn: 9781605587998. doi: 10.1145/1772690.1772751.

[22] L. Scissors, M. Burke, and S. Wengrovitz, "What's in a like? attitudes and behaviors around receiving likes on facebook," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work amp; Social Computing*, ser. CSCW '16, San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 1501–1510, isbn: 9781450335928. doi: 10.1145/2818048.2820066.

[23] S. Seo, J. Kim, S. Kim, J. Kim, and J. Kim, "Semantic hashtag relation classification using co-occurrence word information," *Wirel.*

Pers. Commun., vol. 107, no. 3, pp. 1355–1365, 2019. doi: 10.1007/s11277-018-5745-y.

[24] A. Belhadi, Y. Djenouri, J. C. Lin, C. Zhang, and A. Cano, “Exploring pattern mining algorithms for hashtag retrieval problem,” *IEEE Access*, vol. 8, pp. 10 569–10 583, 2020. doi: 10.1109/ACCESS.2020.2964682.

[25] P. Dooley and B. Božić, “Towards linked data for wikidata revisions and twitter trending hashtags,” in *Towards Linked Data for Wikidata Revisions and Twitter Trending Hashtags*, ser. iiWAS2019, Munich, Germany: Association for Computing Machinery, 2019, pp. 166–175, isbn: 9781450371797. doi: 10.1145/3366030.3366048.

[26] M. Efron, “Hashtag retrieval in a microblogging environment,” in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR ’10, Geneva, Switzerland: Association for Computing Machinery, 2010, pp. 787–788, isbn: 9781450301534. doi: 10.1145/1835449.1835616.

[27] F. Zamberi, N. Adli, N. Hussin, and M. Ahmad, “Information retrieval via social media,” *International Journal of Academic Research in Business and Social Sciences*, vol. 8, pp. 1375–1381, Jan. 2018. doi: 10.6007/IJARBSS/v8-i12/5239.

[28] O. Tsur and A. Rappoport, “What’s in a hashtag? content based prediction of the spread of ideas in microblogging communities,” in *WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, May 2012, pp. 643–652. doi: 10.1145/2124295.2124320.

[29] S. Fedushko, Y. Syerov, and S. Kolos, “Hashtag as way of archiving and distributing information on the internet,” in *Modern Machine Learning Technologies, Workshop Proceedings of the 8th International Conference on ”Mathematics. Information Technologies. Education”*, M. Emmerich, V. Lytvyn, I. Yevseyeva, V. B. Fernandes, D. Dosyn, and V. Vysotska, Eds., ser. CEUR Workshop Proceedings, vol. 2386, 2019, pp. 274–286.

[30] S. Sedhain, S. Sanner, D. Braziunas, L. Xie, and J. Christensen, “Social collaborative filtering for cold-start recommendations,” in *Proceedings of the 8th ACM Conference on Recommender Systems*, ser. RecSys ’14, Foster City, Silicon Valley, California, USA: Association for Computing Machinery, 2014, pp. 345–348, isbn: 9781450326681. doi: 10.1145/2645710.2645772. [Online]. Available: https://doi.org/10.1145/2645710.2645772.

[31] W.-X. Zhou, D. Sornette, R. A. Hill, and R. I. M. Dunbar, “Discrete hierarchical organization of social group sizes,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 272, no. 1561, pp. 439–444, Feb. 2005, issn: 1471-2954. doi: 10.1098/rspb.2004.2970.

[32] P. Lorenz-Spreen, F. Wolf, J. Braun, G. Ghoshal, N. Djurdjevac-Conrad, and P. Hövel, “Tracking online topics over time: Understanding dynamic hashtag communities,” *Computational Social Networks*, vol. 5, pp. 5–9, 2018. doi: 10.1186/s40649-018-0058-6.

[33] R. Kay and S. Loverock, “Assessing emotions related to learning new software: The computer emotion scale,” *Computers in Human Behavior*, vol. 24, pp. 1605–1623, Jul. 2008. doi: 10.1016/j.chb.2007.06.002.

[34] A. Bangor, P. Kortum, and J. Miller, “Determining what individual sus scores mean: Adding an adjective rating scale,” 3, vol. 4, Bloomingdale, IL: Usability Professionals’ Association, May 2009, pp. 114–123.

[35] I. Jakovljevic, “Codis survey tool,” doi: 10.5281/zenodo.5345121.

Developing Situational Awareness from Blogosphere: An Australian Case Study

Mainuddin Shaik[1], Muhammad Nihal Hussain[2], Zachary Stine[1], and Nitin Agarwal[1]
[1]COSMOS Center, [2]Equifax Inc.
[1]Little Rock, United States, [2]Alpharetta, United States
e-mails: [1]{mxshaik, zkstine, nxagarwal}@ualr.edu, [2]nihalhussain.muhammad@equifax.com

***Abstract*— Analyzing topics of interest from online discourse can be challenging. One well-known approach is to conduct topic modeling to study the topics of interest. In this paper, we use a multi-method analytical framework to analyze topics in addition to characterizing their influence. We analyze 20,066 blog posts and 10,113 comments from July 2019 through December 2020 that deal with diplomacy, defense, trade, and election related topics surrounding Australia and China. Our results show that COVID-19 discourse absorbs much of the attention of bloggers during the time period considered, even though no COVID-related keywords were incorporated in the data collection. Our findings suggest that a topic can be influential even when it is not trending and vice-versa. It also showed that Australian bloggers were dominant and discussed the topics of interest compared to Russian and US bloggers. The Australian blogosphere simultaneously discussed climate change along with defense related topics, and they prefer to give attention to long-term topics over short-term topics. Finally, popular election topics in Australian blogosphere tend to have a more analytical tone in the text.**

Keywords-COVID-19; Australia; China; trade war; election; South China Sea; Indo pacific; influence.

I. INTRODUCTION

The increasing proliferation of social media provides tremendous opportunities for gaining situational awareness to assist with strategic policy making, particularly in security, defense, and foreign policy. However, social media data are often riddled with challenges such as high volume and velocity, noisy data, missing data, and incomplete data. Analysis of trending topics, influential topics, and key influencers, afford computational, systematic, and rigorous methods to study big social media data and surmount these challenges [1][2]. The benefits of identifying trending and influential content from social media are already evident: they are often discourse-movers and key indicators that are worthy of additional attention. Identifying these contents can help stakeholders to better understand the key issues and detect, analyze, and monitor new trends by focusing on content generators. Significant previous work has offered solutions but primarily focused on developing single method analysis to solve the issue of accessing relevant information [3][4]. However, by considering the nature of real-world online discourse, we argue that usage of a single method or a tool is limited and instead a multi-method integrated analytical framework can yield multifaceted information well suited for the complex objective oriented studies [5][6].

In this paper, we therefore present a framework to build an interchangeable multimethod analysis with available methods and computational tools to study relevant information or topics of interest from the targeted online discourse. We turn to the blogosphere discourse related to defense, trade, diplomacy, and elections surrounding Australia and China between July 1, 2019, to December 31, 2020 as a case study due to growing diplomatic dispute with USA at the center of the narratives [16][17]. Our research contributes to the growing body of literature on social media analytics to access core, relevant, and influential information from trending online discourse. This illustrative study offers guidelines and strategies for policy makers to identify and monitor key actors from the influential discourse and address the key concerns that may need attention.

The subsequent sections in this paper are arranged as follows. First in section II, we review prior work on natural language processing, influential blogposts, dynamic network analysis and explain how they can be employed in a complementary way tailored toward multiple facets of situational awareness information. In section III, following an overview of the data collection, we summarize how each component of our proposed pipeline maps to our blog dataset. Finally, we present the results of our integrated analysis, assessing topics and their influence trends, the narrative they spread, and their core influential bloggers in the overall online discourse in section IV. With our findings, we concretely demonstrate how our integrated multimethod analytical framework can refine trending influential topics and bloggers through gathering and triangulation of different streams of analysis. We conclude with a discussion of enhancing methodological frameworks for situational awareness information from blogs in section V.

II. LITERATURE REVIEW

Ever-growing scholarship has studied how to monitor trending information and their influential operators with notable success. In this section, we present a review of this literature, highlighting methodologies and gaps in the current state of research on blog analysis.

Latent Dirichlet Allocation (LDA) is a probabilistic generative model that is used to extract topics from a corpus of text documents. While other distributional semantic models can be used to learn latent features from text, the latent features, or topics, that LDA learns tend to be more interpretable by humans than other methods. The topics are

 ISBN: 978-1-61208-899-0

represented as probability distributions over words, since the topics that LDA learns are in the form of probability distributions over the vocabulary of a corpus, reflecting the probability of each word within each topic [7]. Other topic models have been proposed that capture how the topics are spread out in corpus by extracting key information relevant to time of interest [8].

However, even if the dominant or main topics are trending in the blogosphere during a certain period it is not necessary that they are also influential. Researchers have formulated influence as a factor of four parameters: *Recognition, Activity generation, Novelty,* and *Eloquence* [9]. A topic may become influential when a blogger posts content that attracts participation and recognition. Identification of influential topics is therefore important to our multimethod approach.

Finally, we are also interested in the bloggers that posted influential content and asking if they share the same interest in different topics. Influential bloggers prefer to join forces and work in closely interconnected networks. They build their own interest groups, where they gather support, initiate socio-political discussions, and spread awareness [1]. We leverage previously published work to identify a network of core users from online information operations using Organization Risk Analyzer (ORA) [10]. ORA enables analysis of multi-modal networks to shed light on the core influential bloggers that participated in multiple topics.

In this paper, we demonstrate that our multimethod analysis pipeline can provide comprehensive and relevant insights on influential trending topics and users from the topics of interest in the blogosphere. Past studies have shown how accessing situational awareness information requires a multimethod analysis model for comprehensive insights [11][12]. Our proposed framework consists of a multimethod analysis using topic modeling and influence analysis in parallel to discover key insights from blog data.

III. METHODOLOGY

In this section, we describe the data used for the study and our research methodology.

A. Data Collection

We collected the data using a set of keywords provided by subject matter experts at the University of Sydney and Australian Department of Defense. Relevant keywords were identified by an iterative cycle of studying coverage related to diplomatic, defense, trade, and election news between Australia and China. News coverage was reviewed from a sample of news articles published between January 1 and April 30 of 2020, which constituted the first step of the search string. Further reviews of the coverage provided additional keywords to enhance the inclusiveness. The finalized set of keywords used in the study were *'Australia', 'China', 'South China Sea', 'Military Region', 'Fight', 'Tensions', 'Election', 'India' 'Road Belt Initiative', 'Thai', 'Defence' 'Defense' 'Indo pacific', 'Asia Pacific'*.

We then deployed an in-house custom web crawler that invokes Google Custom Search API to collect the URLs that have the required keywords in their articles [13]. Since this study focuses on blogs, we initially targeted the sites known to host blogs, such as blogpost.com, wordpress.com and livejournal.com. We then deployed the crawler on blogrolls obtained from these platforms to target sites that are outside the initial searched domains.

We collected data for 18 months starting from July 1, 2019, to December 31, 2020, and curated 20,066 relevant blog posts in total. We found that the data set yielded many COVID-19 related articles, which in turn skewed away our analysis from the topic of interest. Therefore, we divided the data into two groups to study the effect of COVID-19 over the blog discourse separately: 1) a dataset including all 20,066 blog posts (i.e., All data) and 2) a dataset that excludes any COVID-19 related data (i.e., No COVID data), consisting of 10,113 blog posts. We performed a trial-and-error method to identify required COVID-19 related terms in 'All data' and used the following keywords *'+Australia -(covid covid-19 coronavirus pandemic vaccine)'* to exclude the related posts.

Table I shows the high-level statistics. The data is available for public use through the BlogTracker application [14][15].

TABLE I. BLOG DATA STATISTICS

Data elements	All data	No COVID data
Total posts	20,066	10,113
Total blog domains	679	344
Total blog authors	4,217	2,494
Comments	417,050	84,798

B. Topic modeling

In order to identify the primary themes of the blog posts, we trained two LDA models. One model was trained on the All data set and a second model was trained on the No COVID data set. Prior to model training, basic preprocessing was performed on the text including the removal of stop-words and punctuation. Each model was specified to have five topics. After training, we interpreted the five topics in each data set based on the highest-probability words in each and by reading exemplar blog posts for each topic.

C. Influence

Influence of a blogpost can be visualized as an influence graph or i-graph, where influence flows through the nodes for any given blogpost and each node of a single blogpost in an i-graph is characterized by four variables: 1) Recognition (number of in-links, represented by ι), 2) Activity generation (number of comments, represented by γ), 3) Novelty (number of outlinks, represented by θ), and 4) Eloquence (length of the blogpost, represented by λ) [9]. In a directed i-graph, if $f(x)$ denotes the influence flow of a blogpost x

with I being the influence of x. The influence of each blogpost $I(x)$, can be calculated as

$$I(x) = w(\lambda) \times \left(w_{com}\gamma_x + f(x) \right) \quad (1)$$

where $w(\lambda)$ represents the weight of the blogpost length and $w_{com}\gamma_x$ represents the weight of the total number of comments from blogpost x. $f(x)$ is defined as

$$f(x) = w_{in} \sum_{m=1}^{|\iota|} I(x_m) - w_{out} \sum_{n=1}^{|\theta|} I(x_n) \quad (2)$$

where w_{in} and w_{out} are the weights of the incoming and outgoing influence, x_m is the number of incoming links to blogpost x, and x_n is the number of outgoing links of x. The total number of inlinks and outlinks of x are denoted by $|\iota|$ and $|\theta|$ respectively.

As defined in (2), $f(x)$ measures the difference between total incoming links and total outgoing links of blogpost x. We calculated these four variables from the collected blogposts and calculated the influence of each post as defined in (1). We then mapped the influence scores to the respective topics to generate influence trends for each topic. To do so, we multiplied the influence score $I(x)$ of each blogpost with the topic density (p) score of each blogpost p(topic|blogpost).

$$m(x) = I(x) \times p(topic|blogpost) \quad (3)$$

where, $m(x)$ is the product value for each blogpost. Next, we calculated the mean $m(x, t)$ of each topic for all posts during a given time-period (t) to get the average influence of each topic.

In the next section, we explain our results in the following sequence:

- Conduct topic modeling to identify themes and trends of the discourse in blogosphere qualitatively.
- Assess influence of the identified topics over the period.
- Extracted network of closely connected bloggers from the identified topics and their influence scores to evaluate their interest in various topics.

IV. RESULTS

The highest probability words in the five topics trained on All Data can be seen in Table II. Topics 0, 1, and 2 primarily focuses on Australia and China, whereas topics 3 and 4 focuses on COVID-19 and USA politics, accounting for 44% of the posts. Similarly, the highest probability words from the five topics learned from the No COVID dataset are given in Table III. Our following results focus only on topics relevant to our interests.

A. *South China Sea, China, Australia, USA, Defense tension*

Discourse in this theme focuses on topics such as the Quad alliance [16] and New Silk Road [17]. Content from topic 2 is a crossover between Quad alliance and climate change issues. It discusses Quad by embedding it in a broader context related to Pacific Island foreign policies. Many of these posts simply described the ongoing activities in the South China Sea and Indo-Pacific region. Others were divided based on their stance and opinions on China's actions in the South China Sea and Indo-Pacific region. We observed there were five types of blogposts in this theme: 1) Pro-Australia 2) Pro-China 3) Military and Technology 4) Critical towards Australia government 5) News carrier. Topic 1 from 'All Data' and topic 3 from 'No COVID data' maintained a low profile in 2019 (see Figures 3 and 4) and were not very influential even though these topics were trending in 2019 (see Figures 1 and 2). In other words, bloggers may be actively posting or publishing blogs on certain topics but may not gain enough traction to trigger the required influence parameters [9]. We observe the opposite scenario with topic 2 in 2019, which further indicates that a topic can be influential without trending.

TABLE II. LDA TOPICS FROM 'ALL DATA'

Topic 1. china, chinese, military, india, war, states, security, power, trump, united
Topic 4. covid, vaccine, health, coronavirus, virus, cases, pandemic, public, deaths, care
Topic 2. climate, global, market, trade, energy, economics, china, change, australian
Topic 3. trump, see, america, great, biden, news, get, html, video, know
Topic 0. election, party, vote, policy, bank, zealand, voters, per, labor, think

TABLE III. LDA TOPICS FROM 'NO COVID DATA'

Topic 3. china, chinese, states, war, government, military, power, united, may, use
Topic 0. women, get, much, life, well, made, see, way, back, know
Topic 1. government, australian, china, state, climate, media, per, much, news, change
Topic 2. china, australian, india, south, water, sea, government, pacific, country, fire
Topic 4. party, government, election, vote, labor, australian, votes, parties, preferences, candidates

Most of the influential content were produced in May, November, and December of 2020. In May 2020, blogposts extensively published about the Hong Kong protests regarding security law and raised concerns about a possible exodus of Hongkongers to Australia, New Zealand, Malaysia, or Taiwan. In November 2020, influential blogposts mocked the Australian Elite Special Forces in the killing of Afghans. We noticed much of this influential content was posted by Russian bloggers. In December 2020, influential content expressed disdain for participants in the Quad exercise such as Japan and Australia and suggested they will denigrate the relation with Pacific Islanders.

TABLE IV. MAPPING TOPIC THEMES BETWEEN 'ALL DATA' AND 'NO COVID DATA'

Theme of the topic	All data	No COVID data	Sample blogposts
A. South China Sea, China, Australia, USA, Defense tension	Topic 1	Topic 2, Topic 3	*'Beijing's Hong Kong plans may lead to an exodus, and Australia must be ready'. 'huh, fuck the quad'. 'The massive, systemic and grave crimes committed over the years by fighters of Australian elite'. 'Singapore-Australia exercise activities involves Singapore's F-15SGs'.*
B. Climate change, Economy, and Trade	Topic 2	Topic 1, Topic 2	*'Too much fuel causes extreme bush fires, not climate change'.*
C. Federal and Regional elections in Australia	Topic 0	Topic 4	*'Scott Morrison is up two on approval to 66% and down two on disapproval to 30%, while Anthony Albanese is up one to 44% and up two to 41%, with Morrison's lead as preferred prime minister out from 58-29 to 60-28'*

B. *Climate change, Economy and Trade*

Most of the content in this theme were produced during Australia's bushfire incident in January 2020 and trade tension from April through December of 2020. We found blogposts on bushfires to engage in climate change denial and conspiracy theories. Influential blogposts in both periods were largely limited to climate change and renewable energies, discussed by the Australian blogosphere even though Australia experienced a recession for the first time in 2020 after nearly thirty years [18] and experienced a trade war with China [19]. The only influential conversation on trade war occurred in topic 1 in December 2020, and content during this period criticized the Australian government for its trade relations towards China. It appeared the online public-sphere preferred discussion on climate change over the economy and trade war. This reinforces findings from previous studies that the Australian blogosphere considers certain topics as long-term and short-term issues [20]. In this case, climate change is a long-term issue, and the economy and trade war are short-term issues.

C. *Federal and Regional elections in Australia*

The Australian election theme appeared to trend low compared to the rest of the topics (see Figures 1 and 2). However, it was influential in both 2019 and 2020. In 2019, the discourse was primarily about the labor party defeat in the May 2019 election including discussions about political studies on the election. The aim of these discussions was to identify why the Labor party failed to appeal to the Australian public with extensive analysis. A possible explanation for these influential topics is that readers prefer comprehensive analysis from blogs. In November 2020, influential content (see Figures 3 and 4) discussed the Queensland election and included articles that compared the candidates from various parties based on the policies they support and the likelihood of their win. We also found blogposts with live election updates were less influential than the blogposts that provide analysis on election news.

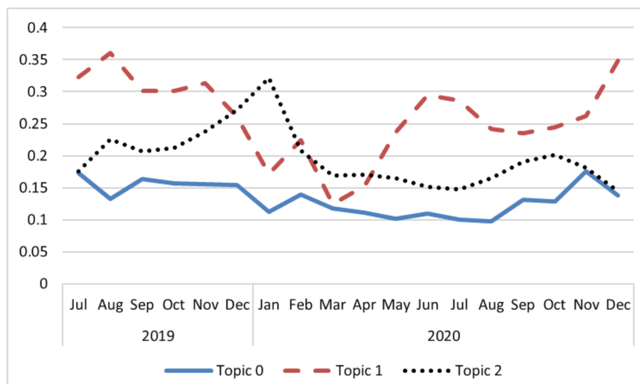

Figure 1. Blogpost topics trends from 'All data'.

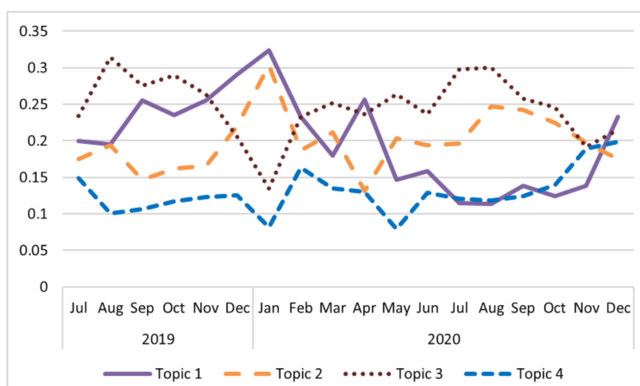

Figure 2. Blogpost topics trends from 'No COVID data'.

 ISBN: 978-1-61208-899-0

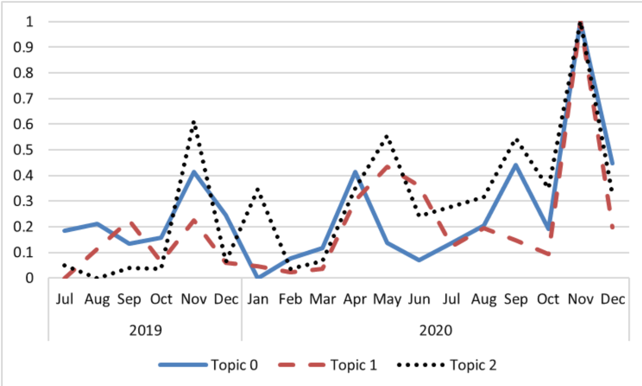

Figure 3. Influence of blogpost topics from ‘All data’.

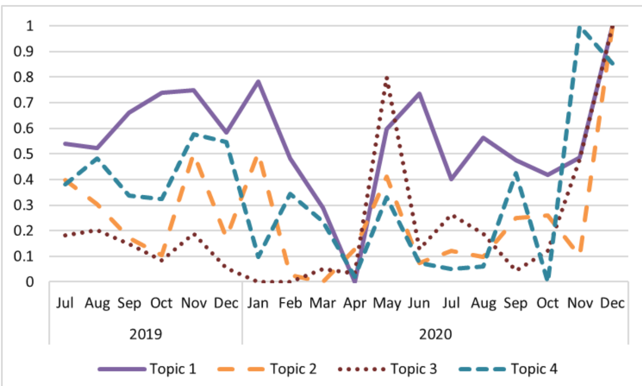

Figure 4. Influence of blogpost topics from ‘No COVID data’.

D. *Topics and Blogger Network*

A network analysis of the top bloggers and their usage of topics provides insights on which bloggers share interest in the same topics. Figure 5 depicts the clusters of bloggers in each topic and network shows the bloggers who blogged in five topics from ‘All data’. Similarly, Figure 6 network shows the bloggers and the topics they belong to in ‘No COVID data’. We found both topics 1 and 2 from ‘All data’ and topics 3 and 1 from ‘No COVID data’ appeared to have a large number of bloggers in their respective networks, indicating a good number of bloggers in both data discussed diplomacy, defense, and trade related topics. However, the cluster of bloggers in the election’s topic were comparatively less in both data groups aligning with our findings related to theme ‘c’ that not many bloggers were active in the election related topic and therefore produced a smaller number of blogposts compared to other topics.

Networks from both groups also show a cluster of bloggers in the center of the network who contributed to multiple topics, indicating that the core bloggers actively participated in all five topics. We wanted to examine the influence of these core bloggers and extracted them by folding the networks based on top valued links. In Figure 7(a) network shows two bloggers contributed to multiple topics from ‘All data’ based on their edge color. Black nodes represent bloggers, while orange edges represent four topics and red edges represent five topics. Similarly, 7(b) from ‘No COVID data’ shows the connection between two bloggers in three topics, four topics and five topics, where the edges is blue, green and red.

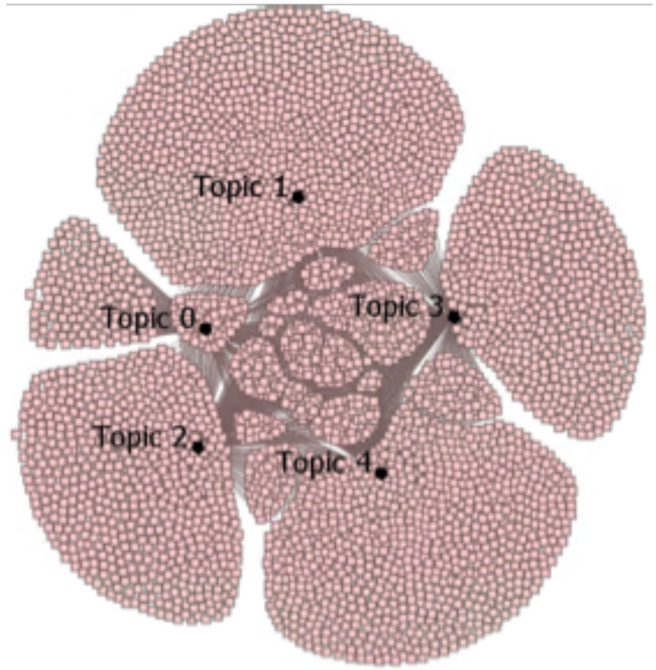

Figure 5. Topics and Blogger network from ‘All data’

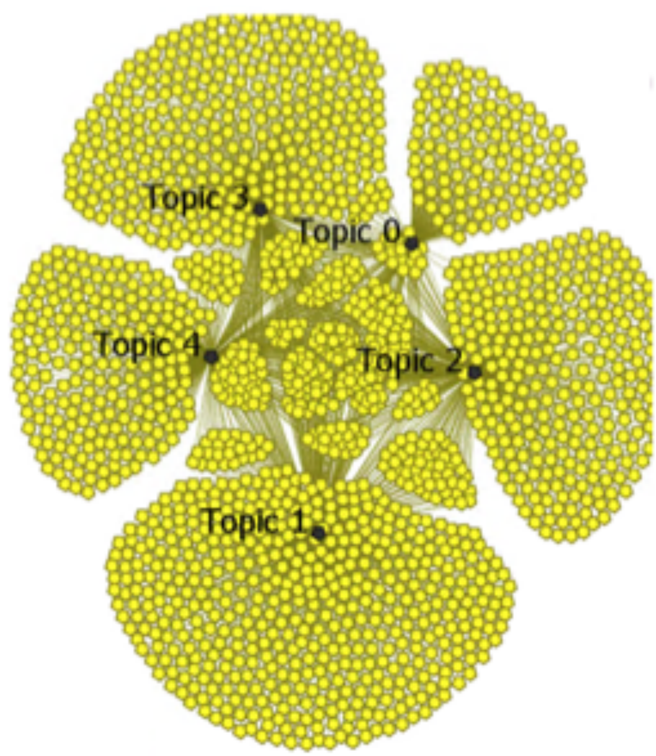

Figure 6. Topics and Blogger network from ‘No COVID data’

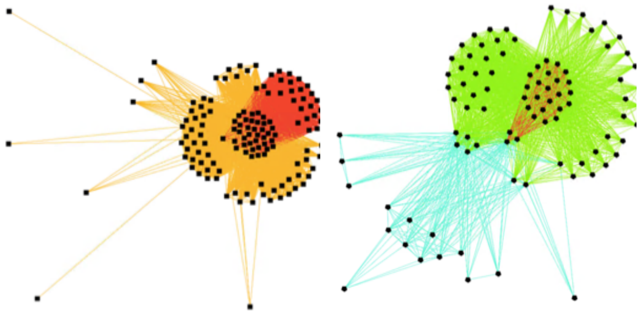

Figure 7. Folded Topics and Blogger network from 'All data'(a) and 'No COVID data' (b)

In Table V, we present core bloggers and their cumulative influence score in each group.

TABLE V. CORE BLOGGERS AND THEIR INFLUENCE SCORE

'All Data'		
Blog	*Blogger*	*Influence score*
catallaxyfiles.com	Sinclair Davidson	11100.6
catallaxyfiles.com	currencylad	5260.5
theaimn.com	The AIM Network	4151.4
quadrant.org.au	quadrant	1566.5
catallaxyfiles.com	Guest Author	1536.6
theaimn.com	Dr Binoy Kampmark	1488.3
pngattitude.com	Keith Jackson	198.6
crikey.com.au	Charlie Lewis	6.6
'No COVID data'		
Blog	*Blogger*	*Influence score*
zerohedge.com	Tyler Durden	23678.7
catallaxyfiles.com	currencylad	2560.5
theaimn.com	The AIM Network	1925
crikey.com.au	John Quiggin	753.9
quadrant.org.au	quadrant	613.5
theaimn.com	Dr Binoy Kampmark	553.8

We found highly influential core bloggers in both groups who participated in multiple topics belongs to Australian blogosphere.

V. CONCLUSION & FUTURE WORK

In this research, our multi-faceted integrated analysis offers a multidimensional view on blog discourse surrounding Australian defense related topics. Our analysis provides several useful insights. We determined the Australian bloggers were the major discourse movers and produced trending, influential content on defense and climate change related topics. These bloggers' opinions and stance on defense related topics were mostly pro-Australia and disapproved of China's military activities in the South China Sea. Additionally, we identified climate change content largely in denial of climate change and espousing conspiracy theories around renewable energy.

We observed that, even though a topic was trending at a certain period, it was not necessarily influential at the same time. We also captured the behavior pattern of the Australian blogosphere during election that influential content tends to have analytical tone. Additionally, we found that COVID-19 was influential only for a short period of time in the blogosphere despite it being a global health crisis. Further, COVID-19 topic did not gain much traction in the blog discourse compared to the defense and climate change topics. This may indicate that the blogosphere considers COVID-19 as a short-term issue and defense and climate change topics as long-term issues. This observation can be investigated in future work to study bloggers' preferences. Addition, exploration can be done to see whether there is harmonization over transcontinental social communities, in terms of "igniters", "followers", and "passive consumers" that can help various governmental or security agencies to monitor key influencers. Further, an in-depth analysis of the relationship between topical trends and their influence is warranted based on the findings.

ACKNOWLEDGMENT

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

REFERENCES

[1] I. Kayes, X. Qian, J. Skvoretz, and A. Iamnitchi, "How influentialare you: Detecting influential bloggers in a blogging community," in *International conference on social informatics*. Springer, 2012, pp. 29–42.

[2] L. Dahlberg, "Re-constructing digital democracy: An outline of four 'positions'," *New media & society*, vol. 13, no. 6, pp. 855–872, 2011, number: 6 Publisher: Sage Publications Sage UK: London, England.

[3] V.-A. Nguyen, et al, "Modeling topic control to detect influence in conversations using nonparametric topic models," *Machine Learning*, vol. 95, no. 3, pp. 381–421, 2014, number: 3 Publisher: Springer.

[4] Z. Yang, C. Wang, F. Zhang, Y. Zhang, and H. Zhang, "Emerging rumoridentification for social media with hot topic detection," in *2015 12th Web Information System and Application Conference (WISA)*. IEEE, 2015, pp. 53–58.

[5] J. Uyheng and K. M. Carley, "Characterizing bot networks on Twitter: An empirical analysis of contentious issues in the Asia-Pacific," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 2019, pp. 153–162.

[6] M. W. DiStaso and D. S. Bortree, "Multi-method analysis of trans- parency in social media practices: Survey, interviews and content analy-sis," *Public Relations Review*, vol. 38, no. 3, pp. 511–514, 2012, number:3 Publisher: Elsevier.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation,"*the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003, publisher: JMLR. org.

[8] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 113–120.

[9] N. Agarwal, H. Liu, L. Tang, and S. Y. Philip, "Modeling bloggerinfluence in a community," *Social Network Analysis and Mining*, vol. 2, no. 2, pp. 139–162, 2012, number: 2 Publisher: Springer.

[10] K. M. Carley, J. Diesner, J. Reminga, and M. Tsvetovat, "Toward an interoperable dynamic network analysis toolkit," *Decision Support Systems*, vol. 43, no. 4, pp. 1324–1347, 2007, number: 4 Publisher: Elsevier.

[11] S. Hofmann, D. Beverungen, M. Räckers, and J. Becker, "What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook," *Government information quarterly*, vol. 30, no. 4, pp. 387–396, 2013, number: 4 Publisher: Elsevier.

[12] M. Benigni, K. Joseph, and K. M. Carley, "Mining online communities to inform strategic messaging: practical methods to identify community- level insights," *Computational and Mathematical Organization Theory*, vol. 24, no. 2, pp. 224–242, 2018, number: 2 Publisher: Springer.

[13] M. Allauddin and F. Azam, "Service crawling using google custom search api," *International Journal of Computer Applications*, vol. 34, no. 7, p. 2011, 2011, number: 7 Publisher: Citeseer.

[14] A. Obadimu, M. N. Hussain, and N. Agarwal, "Blog data analytics usingblogtrackers," in *Big Data and Social Media Analytics*. Springer, 2021, pp. 113–125.

[15] Blogtrackers, "Analyze anything about blogs," University of Arkansas at Little Rock, Jan. 2021. [Online]. Available: https://btracker.host.ualr.edu , [retrieved: Jan, 2021].

[16] "The Quad in the Indo-Pacific: What to Know," Jun. 2021. [Online]. Available: https://www.cfr.org/in-brief/quad-indo-pacific-what-know, May. 27, 2021. [retrieved: June, 2021].

[17] "China's Massive Belt and Road Initiative," Jun. 2021. [Online]. Available: https://www.cfr.org/backgrounder/chinas-massive-belt-and-road-initiative, Jan. 28, 2020. [retrieved: June, 2021].

[18] "Australia Enters Its First Recession in Nearly 30 Years - The New York Times," July. 2021. [Online]. Available: https://www.nytimes.com/2020/09/02/business/australia-recession.html, Sep. 02, 2020. [retrieved: July, 2021].

[19] S. Grant, S. Dziedzic, and B. Xiao, "Is iron ore our secret weapon? Your big questions about China trade war answered," *ABC News*, Dec. 2020. [Online]. Available: https://www.abc.net.au/news/2020-12- 10/china-australia-trade-war-your-questions-answered/12971434. [retrieved: July, 2021].

[20] L. Kirchhoff, T. Nicolai, A. Bruns, and T. Highfield, "Monitoring the Australian blogosphere through the 2007 Australian federal election," in*Communication, Creativity and Global Citizenship: Refereed Proceed- ings of the Australian and New Zealand Communication Association Conference 2009*. ANZCA, 2009, pp. 982–1005.