Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems | School of Computing and Information Systems

8-2021

Discovery of mental wellness via social analytics for liveability in an urban city

Kar Way TAN
Singapore Management University, kwtan@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, and the Urban Studies and Planning Commons

Citation

TAN, Kar Way. Discovery of mental wellness via social analytics for liveability in an urban city. (2021). *2021 17th IEEE International Conference on Automation Science and Engineering (CASE): Lyon, France, August 23-27: Proceedings*. 1-6.

Available at: https://ink.library.smu.edu.sg/sis_research/6692

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

Discovery of Mental Wellness via Social Analytics for Liveability in an Urban City

Kar Way Tan [1]

Abstract— **Smart cities, are often perceived as urban areas that use technologies to manage resources, improve economy and enhance community livelihood. In this paper, we share an approach which uses multiple sources of data for evidence-based analysis of the public's views, concerns and sentiments on the topic related to mental wellness. We hope to bring forth a better understanding of the existing concerns of the citizens and available social support. Our study leverages on social sensing via text mining and social network analysis to listen to the voices of the citizens through revealed content from web data sources, such as social media and public forums. By using hybrid data sources, we present the important considerations for mining inherent mental wellness concerns faced by the citizens. The outcome of the analysis includes, both the positive and negative sentiments towards mental wellness and draws relations to national level performance indicators relating to mental wellness. We hope our research could help authorities derive actionable plans for designing health services or public events that bring positive social mixing and happiness by addressing the mental wellness of the residents.**

I. INTRODUCTION

Mental wellness refers to the positive state of mental health, and it involves a high level of mental relaxation of the individual [1]. Mental illness involves the disturbance of the mind to the extent where it affects not only the victims but also their families and communities. At the extreme end, suicide and death due to mental disorder are disturbances in an urban city. Early detection and identification of those at risk of mental illness are important challenges.

Putting it in perspective, intentional self-harm such as suicide is a global issue affecting many regions of the world. According to the World Health Organization, more than 700,000 lives are lost annually due to suicide [2]. It reports that the number of people attempting suicide is higher than the number of suicide deaths. It is reported that the global suicide rate in developed countries had been increasing since 1999 and that by 2010, suicide had become the major reason behind unnatural death [3].

Singapore is no exception. In recent years, increasingly more cases of elderly suicide have been reported in Singapore [4]. Among the younger age group of 10 to 29, suicide is known to be the leading cause of their death [5]. Another study reports that nearly 92% of the working Singaporeans were stressed, and this is much higher than the global average of 84% [6]. This signifies an apparent need in the management of mental health in Singapore.

For over 20 years, the Internet has drastically changed how the world communicates, especially with the widespread usage of social media sites. Prior to social media, people were limited in their means to interact with others. Interactions were limited to the people that they knew face-to-face. Today, social media sites not only connect people from all over the world, but they also provide platforms for people to express their opinions using text, images, and videos. Social media analytics seeks to gather and analyse data from social networking sites, and it enables discovery of insights such as users' personalities. The works in [7], [8], [9] show that deep content analysis on unstructured data can reveal each person's choice of vocabulary and syntax, and it can reveal important insights about the individual's personality, emotions, fears, hopes, and motivations. The work in [10] outlines methods and findings to use community-specific language features to discover personal values among the social media users. Two other works in [11] and [12] show social media can reveal human behaviour via unstructured text analysis and better understand sentiments in difficult times, e.g., during the pandemic period.

There have been multiple works based on social media analytics and big data analysis in identifying individuals who are at risk of depression and suicide [13], [14]. The work by [15] proposed multi-model detection method to reveal the underlying online behaviors beneath the mental health problems. The 2017 incident involving Facebook revealed that the company's algorithms could identify young people during moments of vulnerability by analysing their word usage (e.g. 'stressed', 'worthless', 'insecure', 'anxious', 'useless', 'overwhelmed' among others) and photos. The algorithms can contribute to the detection and identification of online users at risk of mental illness.

In this research, we focused on discovering, both mental wellness and illness, based on revealed emotions on public platforms, such as social media and forum. We proposed a method which would first discover the communities which mentioned mental wellness or illness on Twitter. Then we identified the topics (both positive and negatives ones) to discover the support, concerns and issues faced by the users. In order to gain deeper insights into the support offered, and concerns faced in localized context, we added a second dimension to analyse data from Reddit (a forum). The posts on Reddit are typically longer and provided more details compared to tweets which have a limited length for each post. We illustrated our proposed approach using Singapore as the case study, focusing on the youth (above 15) and working adults up to retiring age (currently at age 63). Singapore, being a fast-paced and cosmopolitan city with pervasive use of social media among the case group, we believe the

[1]Kar Way Tan is with School of Computing and Information Systems, Singapore Management University, kwtan@smu.edu.sg

dataset is suitable to provide insights to our study. From the results of both the analyses based on Twitter and Reddit data, we then mapped the findings into an Empathy map that summarized the pains and gains of social media users with respect to mental wellness. As part of our analysis, we also drew relationships between the observations on social media and the city's healthcare performance indicator, such as the admission rate to psychiatric hospitals.

Our main contributions are two-fold. Firstly, we proposed an evidence-based approach and a structured analysis method using empathy map for discovering support and concerns related to mental wellness, based on social listening. Secondly, we provided insight to national trends in metrics related to mental health. Specifically, we would be using the admissions to psychiatric hospitals as the metric. We envisioned that these contributions would allow policy-makers to devise policies and design activities to promote liveability in an urbanized environment.

The rest of the paper is structured as follows: Section II provides the objectives and research questions to our study. Section III describes the data we have for our analysis. Section IV describes our proposed approach and results from the analyses. Section V provides additional inferences and discussions based on our study. Finally, we draw our conclusion in Section VI.

II. Problem Definition

We begin our analysis on the impact of mental wellness in the country by using the healthcare datasets available on the Government's Open Data Platform. We used the dataset titled 'Death by cause and age group and gender' [16] to analyse the mortality data for the reasons of death in Singapore. In particular, the International Classification of Disease (ICD) code *V01-Y89* refers to *External causes of morbidity and mortality* and detailed code *X60-X84* refers to mortality related to *intentional self-harm*. A visual representation of the data is plotted on a stacked bar chart shown in Figure 1.

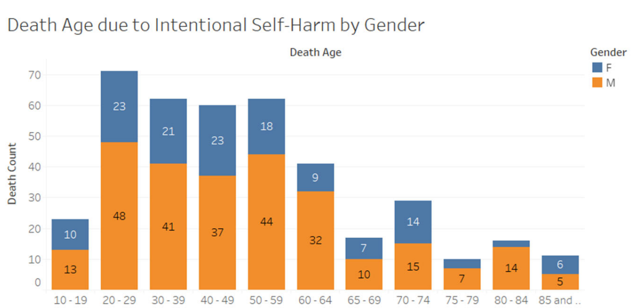

Fig. 1. Death by Intentional Self-harm based on data for year 2019

The statistics revealed that the youth, young to middle-aged adults are especially susceptible to death relating to self-harm. We seek to understand the reasons by social listening on platforms such as Twitter and Forum to provide insights to issues faced and the support available for mental wellness. We believe that the active users on social media are similar to the profile of the target age-group who are of higher risk.

Our key research questions are as follows:

1) What form of support mechanisms and concerns related to mental wellness can be revealed by active users on social media?
2) How do social network analysis and text mining on public platforms (e.g., social media, forums) provide trends that correlate with the macro-level mental wellness metrics (e.g., admission to psychiatric hospitals)?

III. Data

We used data collected from two public sources, namely, Twitter and Reddit (an online Forum). To collect data from Twitter and Reddit, we first compiled a list of standard keywords which were relevant to mental wellness. Next, we added single words, bigrams or trigrams that referred to mental wellness or mental health organizations in the country. In addition, as the locals in Singapore, at times use language constructs which is a hybrid of English and native languages such as Malay and Chinese, there exists Singaporean's colloquial phrases which are frequently found on social media posts. We also considered the localized phrases to better serve our search algorithm for gathering the posts related to mental wellness.

Table I shows the keywords (preprocessed to lowercase) used to filter and collect data from Twitter and Reddit. A few noticeable customized additions to the local language are 'imh', 'samaritans of singapore' and 'no money'. The first two terms refer to the Institute of Mental Health and a non-profit organization in the country and the last term shows an example of local colloquial used to mean lack of money.

TABLE I
Terms Related to Mental Wellness

Type	**Terms**
Indicative of Well-Being	love for oneself, healthy thoughts, self actualization, spirituality, self care, positive mindset, contentment, good social relationships, serenity, practice care, warmth, peace, meditation, sanity, able to cope, stress management, able to manage emotions, resilience, treat yourself, stress-free, articulating thoughts, coping mechanism
Neutral	reach out, emotional regulations, samaritans of singapore, therapy, get help, anti-depressant, counsellor, understanding surroundings
Indicative of Distress	stress, despair, burden, no money, bullying, pressure, worthless, depress, mental tension, burnout, anxiety, too much work, anger, struggle, mental strain, emptiness, worry, bipolar, overworked, insomnia, apprehension, social anxiety, sleep deprivation, dissatisfaction, insecure, frustration, depression, distress, unsound mind, mentally unwell, crazy, disorder, tired, suicide, panic, mental illness, pressuring environment, stigma, depressed, hyperventilation, imh, messy, sad, sos, helpless, insane, lonely, dejected, die, hyperventilate, mood swings, confusion

In the following table, i.e., Table II we summarize the data we acquired, the source and analyses applied to the dataset.

IV. Approach

The overview of our evidence-based approach is depicted in Figure 2. To address the research questions, the first part of the approach covers opinions from a broader spectrum,

TABLE II

DATA SOURCES AND THE ASSOCIATED ANALYSES PERFORMED

Data Sources	Analyses
Twitter geolocation Singapore	1. Network Analysis 2. Topic Modelling 3. Sentiment Analysis
Reddit r/singapore subreddit	1. Topic Modelling 2. Sentiment Analysis with Multi-year Trend Analysis

as represented by the Twitter data. The main objective is to discover the communities using network analysis by analysing the social interactions through 'retweeting', 'replying' and 'following'. Thereafter, drilling down to the few more specific communities and understanding trending themes using Latent Dirichlet Allocation (LDA) [17] topic modeling. Finally, we run sentiment analysis to better understand the polarity of the sentiments in the topics which have been identified.

Twitter imposes a maximum of 280 characters for each tweet, hence the emotions are less well-expressed in these micro-blogs. In contrast, Reddit allows its users to make longer posts, so users can interact with one another via threaded discussions too. Therefore, the objective of using Reddit data is to dwell deeper into the content of the conversations related to mental wellness in the Singapore context. In addition, the standard APIs allow us to collect data from Reddit over a period of a few years. Through a longitudinal dataset from Reddit, we aim to draw insightful correlations to the trends on national indicators relating to mental health.

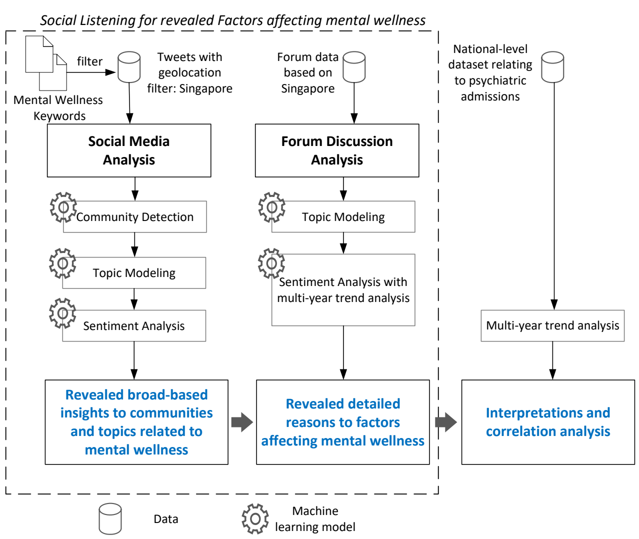

Fig. 2. Approach of our Analysis

A. Broad-based Insights Revealed through Social Analytics

We collected 3878 data points from a snapshot of tweets between 10 17 March 2020, and filtered using keywords related to mental health as per Table I. Then, we applied three analysis methods as follows:

1) Network analysis cum topic modeling - We passed the data through Girvan Newman Algorithm [18] to detect communities. The results showed 24 communities with 20 or more unique users. We then grouped the tweets based on each community and performed LDA Topic Modeling using k=2 (i.e., two main topics discussed in the community). The topics found in each community analysis were then regrouped and generalized to discover the main topics discussed. We found few prominent topics discussed across the communities, they are (a) *COVID-19 fear and concerns*, (b) *Celebrities and K-POP*, (c) *Relationships*, (d) *Work and school(including bullying)*, (e) *Self-care, mental awareness and well-being*, (f) *Politics in other countries*. Some examples of less prominent topics found are related to religion, income, and social issues. There were also few topics which have little meaning and relevance to mental wellness; these will be omitted from our study. The communities were plotted on network graphs; each graph evaluated the influential users within the community. In each community, we then grouped the tweets and applied topic modeling to discover the distribution of words for each topic. An example of topics discovered within the community is shown in Table III. This community provides support to the community at large, promotes mental wellness and self-care on Twitter. Finally, we analysed samples of tweets in each topic to make inferences to the meaning of the topic. In our example, Topic 0 is related to helping the community to be aware of one's state of mental health while topic 1 contains tweets related to help and love for those in need.
2) Pure topic modeling - We used GenSim Python package [19], an open-source library for implementing LDA topic modeling and performing natural language processing, to discover the topics across all the 3878 tweets. For this task, we first filtered extreme words with less than five occurrences. We performed standard text processing steps such as removal of URLs and mentions, replacement of 'n't' with not and removal of the apostrophe such as ''s', ''m', ''d'. The best model found was when k=5 (5 topics). This analysis reflected a slightly different set of topics. The reason is likely due to the community filtering as we kept only the communities that have at least 20 users in the above method. In other words, the topics mined from the network analysis would have dropped out some tweets. The topics discovered in this task are profiled to be *(Topic 0) Bullying, suicide and insecurity*, *(Topic 1) Loneliness and love*, *(Topic 2) COVID-19 concerns*, *(Topic 3) Self-esteem and personal appearance* and *(Topic 4) Workplace, home and family*.
3) Sentiment Analysis - In our next step, we used Valence Aware Dictionary and sEntiment Reasoner (VADER) [20], a lexicon-based text analysis method to evaluate the general sentiment of the tweets. Our analysis revealed that tweets related to mental wellness terms are generally more negative than positive. Using the topics discovered in 2 above, Figure 3 shows the average

sentiment scores of the tweets in each topic. We can see that only Topic 4 reflects a more positive sentiment score compared to the negative ones. We looked into examples of tweets and found a key term ‘resilience’ which attributes to strength and encouragement.

TABLE III
AN EXAMPLE OF TOPICS DISCOVERED THROUGH A COMMUNITY

Topic	Distribution of Words
Topic 0	0.054*”anyone” + 0.037*”never” + 0.037*”true” + 0.037*”lie” + 0.037*”depressed” + 0.036*”fun” + 0.036*”bored” + 0.036*”till” + 0.031*”personal” + 0.031*”nobody”
Topic 1	0.121*”love” + 0.066*”always” + 0.064*”keep” + 0.062*”wan” + 0.060*”mental” + 0.060*”illness” + 0.034*”lookout” + 0.033*”everything” + 0.033*”please”

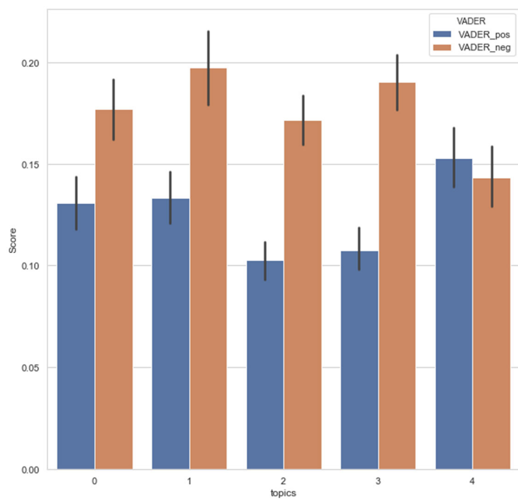

Fig. 3. Sentiment Analysis based on Tweets for each Topic

B. Detailed Insights Revealed through Reddit Forum Data

We collected 12,664 Reddit comments (after dropping posts with very few words) from 2011 to 2020, using subreddit ‘r/singapore’. Our analysis first involved topic modeling using LDA and then sentiment analysis using VADER. Since we could collect comments across the years, we could perform further temporal analysis of the posts, such as trends of unique posts and sentiment scores over the years for each topic (using output from the topic modeling).

There are two steps in this analysis. The methods are described as follows:

1) Topic Modeling – We used GenSim’s LDA package to perform topic modelling across all 12664 comments. We filtered the extreme words with less than 30 occurrences. Similarly, we included text preprocessing steps such as removal of URLs, emoticons, stopwords, punctuations, apostrophes, digits, symbols, replacement of slangs, transformation of negation words and lemmatization. Three distinct topics were observed, they were discussions on (1) Government and Society, (2) Health related and (3) Personal life related, including love and relationships. After this step, each post was assigned to a topic and we moved to the next step.
2) Sentiment Analysis – Similarly, we used VADER to obtain the sentiment polarity and intensity. Next, we performed trend analysis by aggregating the data to obtain the number of unique users and number of posts in each year. We then normalised the sentiments from each topic over each year. The normalized sentiment is computed based on the product of sentiment score from VADER, proportion of unique users and proportion of posts. The normalised sentiment scores were then plotted along the time axis.

The result of the two-step analysis is presented in Figure 4 and 5. We observed that the usage on social media has increased over the years. The *Government and Society* related posts showed increasing sentiments in the initial few years between 2011 and 2015, then sentiment scores dropped after year 2015. For both *Health* and *Personal Life* related topics, sentiment scores showed a downward trend. A deeper dive into individual posts showed that the comments related to *Government and Society* contained discussions on the friendliness of the local mental health facilities and views on the preparedness of the local government in handling the citizen’s needs in this aspect.

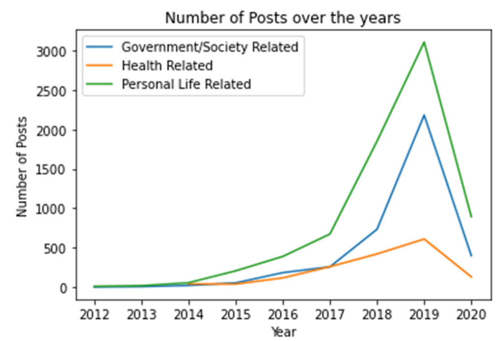

Fig. 4. Number of posts observed on Reddit over years

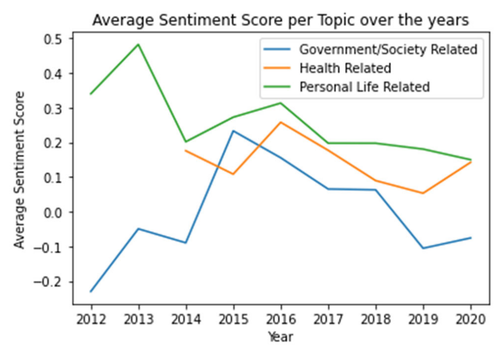

Fig. 5. Sentiment scores based on each Topic over years

V. DISCUSSIONS

In this section, we consolidate the findings from our analyses and present the discussions in relation to our research questions as stated in Section II.

A. Interpretation of Social Media Findings

We draw inspirations from the gaming industry to model personas using an Empathy Map [21]. We put together our analysis and place the insights into a modified version of an Empathy Map as shown in Figure 6 which classified findings into *think and feel, see, hear and do*. The Empathy Map is constructed based on our understanding and inferences of the topics and communities discovered in our approach in Section IV. We also derive insights into the pains and gains of the users based on their involvement on social media platforms by analysing samples of tweets and the words with high coefficients from topic modeling.

With reference to our Empathy map, we address our first research question. Our analyses on Twitter and Reddit data revealed the key topics of what people **think and feel**. People were feeling troubled and emotionally affected by work, life, family and relationship issues. Some were also troubled by bullying and harboured suicidal thoughts. Some others were concerned about the pandemic situation and upset with political issues in other countries. Social media users could **see** the measures taken by the local Government, individuals or organizations in empowering self-care methods and were encouraged to speak up within the related communities. One could **hear** (virtually) the importance of mental wellness and self-care measures to manage an individual's challenges. One also received opinions on personal and health issues, gained perspectives and found ways to slow down the pace and practised mental wellness. Finally, what one could **do** included getting help or providing help to others in the community by receiving or offering thoughts and a listening ear.

Where it used to be, pains in terms of finding avenues of outlet or stigma to talk about mental issues in the local community, people turned to social media, to gain awareness of mental wellness, build a supportive community and provide avenues to seek opinions and perspectives.

We envisioned that our approach provides a way to drawing both positive and negative insights related to mental wellness from social media platforms. It allows citizens to contribute and also benefit from the community, in terms of managing the demands of fast-moving lifestyles in an urban environment.

B. Insights to national level indicators on mental wellness

To address our second research question, we investigated the 'Admissions and Outpatient Attendances' from the Statistics Singapore Open Data Platform [22]. The data contains information about the age group (i.e., 0 to 14 years old, 15 to 64 years old and 65 years and above), type of healthcare facilities (i.e., Acute Hospitals, Psychiatric Hospitals and Community Hospitals), gender and the admission rate in each year starting from 2006 to 2019. For the purpose of this analysis, we used only data relevant to mental health, e.g., admission rates to psychiatric hospitals.

Since our Reddit analysis covers the years 2011 to 2020, we used only data from the overlapping years among the two datasets, i.e., we used only data from 2012 to 2019. Two correlation tests, Pearson (parametric) and Spearman (non-parametric) correlation tests were explored since the data points were limited. We obtained the correlations between the rate of admissions and multi-year trend analysis from the Reddit analysis, and presented the results in Table IV. The column prefix with *User* refers to number of unique users, the column prefixed with *Posts* refers to number of posts and prefixed with *Sentiment* refers to normalised sentiment score.

TABLE IV

REDDIT ANALYSIS AND ITS CORRELATION WITH ADMISSIONS TO PSYCHIATRIC HOSPITAL

Key Measurements	**Pearson Coefficient** (Year 2012 to 2019)	**Spearman Rank Coefficient**
User (Govt)	0.3053	0.2530
User (Health)	-0.7676	-0.1177
User (Personal)	-0.4750	-0.5543
Posts (Govt)	0.2308	0.1928
Posts (Health)	-0.6674	0.0883
Posts (Personal)	-0.5465	-0.3615
Sentiment (Govt)	0.5276	0.6747
Sentiment (Health)	-0.6970	0.0883
Sentiment (Personal)	-0.7542	-0.3374

Based on both correlation analyses, the discussions and sentiments about *Government and Society*, mental wellness initiatives have positive correlation while discussions on *Personal* matters have negative correlation. Discussions on *Health-related* matters seem to suggest negative linear relationship with admission rate but it is non-conclusive when evaluated with non-parametric correlation tool.

We draw insights and potential applications based on relationships found in this analysis. The positive correlation between discussions on the governmental and societal efforts to admission rate suggests that one may get an initial sensing on the concerns and challenges faced by the communities which may lead to higher demand in mental health services on these social media platforms. The negative correlation between discussions on personal matters to admission rate suggests that individuals may voice out unhappiness and concerns or seek help from the community when faced with personal stress. Mental health service providers can also make use of the platforms to propagate mental wellness initiatives and encourage avenues to share thoughts and manage mental wellness at the individual level.

In response to our second research question, our analysis showed potential correlation to national indicators such as psychiatric hospitals. Although we understand this analysis involved only a snapshot of the Twitter data which may not fully imply a longitudinal inference, we hope this finding opens opportunities to better understand the community in an urban city and provide insights into the capacity requirements and planning of mental health services and facilities.

VI. CONCLUSION

In this research, we presented an approach for mining social media data to provide broad views of the communities, topics and sentiment related to mental wellness. We also presented a method for seeking deeper understanding of the mental wellness related conversations by using discussions on forum. With both analyses, we presented, using an empathy map, a better understanding of the pain and gains related to mental wellness through how people think and feel, what they can see, hear and do. In addition, we have also provided a method for drawing insights into national metrics such as admission rate to psychiatric hospital which

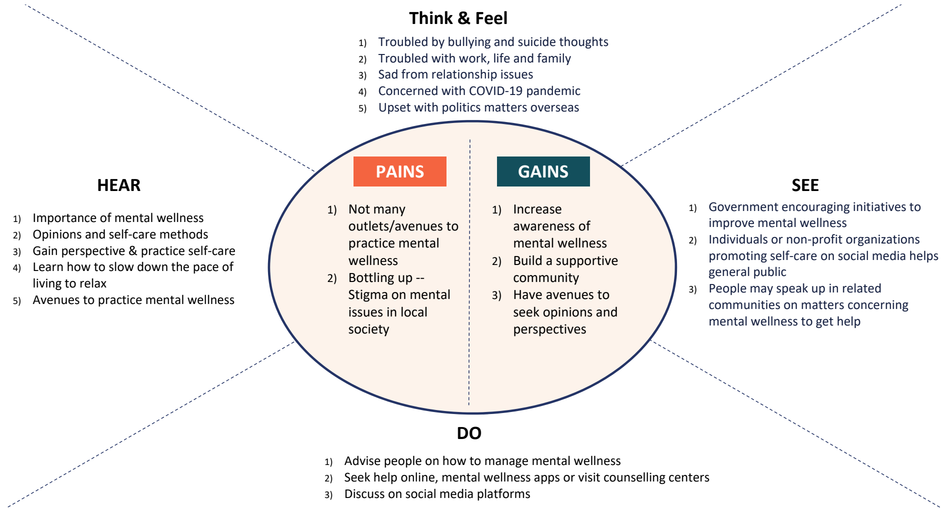

Fig. 6. Empathy map derived from the social media and forum analyses

aids the planning and design for mental health services. We hope our contributions can pave ways to build a stronger community-assisted ecosystem for mental wellness.

For our future work, we can conduct an empirical study focusing on youth and working adults to validate the findings. At the same time, we could collect more data to provide insights over different time periods. In terms of technical tasks, we could use other network and text analytics algorithms to compare and contrast the findings and insights.

ACKNOWLEDGMENT

The author would like to thank the students (Xin Yuan Cheng, De Lin Ong, Janell JE Lee, Mathilda Ng and Su Yee Kok) who contributed parts of data collection and analysis in their course project.

REFERENCES

[1] "What is mental wellness," Singapore Association for Mental Health, https://www.samhealth.org.sg/understanding-mental-health/what-is-mental-wellness, Last accessed on June 15, 2021.

[2] "Mental health and substance use," World Health Organization, https://www.who.int/teams/mental-health-and-substance-use/suicide-data, Last accessed on June 15, 2021.

[3] D. Taylor, "Suicide as a product of modernization," Medium, 2013. [Online], https://medium.com/@dvlan/suicide-as-a-product-of-modernization-776e9823cfaf, Last accessed on June 15, 2021.

[4] R. Rashith, "Number of suicides committed by the elderly hits record high as singapore population ages," The Straits Times, 2018. [Online], https://www.straitstimes.com/singapore/more-than-1-in-3-suicides-committed-by-elderly-as-singapore-population-ages, Last accessed on June 15, 2021.

[5] "Learn about suicide quick facts," Singapore Association for Mental Health, https://www.sos.org.sg/learn-about-suicide/quick-facts, Last accessed on June 15, 2021.

[6] R. W. Neo, "Sleep-deprived singaporean workers among most stressed globally: Survey," TODAYonline, 03 2019. [Online], https://www.todayonline.com/singapore/sleep-deprived-singaporean-workers-among-most-stressed-globally-survey, Last accessed on June 15, 2021.

[7] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," in *Proceedings of Conference on Human Factors in Computing Systems*, 01 2011, pp. 253–262.

[8] S. Gosling, A. Augustine, S. Vazire, N. Holtzman, and S. Gaddis, "Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information," *Cyberpsychology, behavior and social networking*, vol. 14, pp. 483–8, 09 2011.

[9] D. Marengo and C. Montag, "Digital phenotyping of big five personality via facebook data mining: A meta-analysis," *Digital Psychology*, vol. 1, pp. 52–64, 06 2020.

[10] A. Silva, P.-C. Lo, and E. Lim, "On predicting personal values of social media users using community-specific language features and personal value correlation," *ArXiv*, vol. abs/2007.08107, 2020.

[11] M. R. Davahli, W. Karwowski, E. Gutierrez-Franco, K. Fiok, G. Wrobel, R. Taiar, and T. Ahram, "Identification and prediction of human behavior through mining of unstructured textual data," *Symmetry*, vol. 12, 11 2020.

[12] S. Saraff, T. Singh, and R. Biswal, "Coronavirus disease 2019: Exploring media portrayals of public sentiment on funerals using linguistic dimensions," *Frontiers in Psychology*, vol. 12, p. 626638, 02 2021.

[13] M. De Choudhury, S. Counts, and E. Horvitz, "Social media as a measurement tool of depression in populations," in *Proceedings of the 5th Annual ACM Web Science Conference*. Association for Computing Machinery, 2013, pp. 47–56.

[14] X. Yang, R. McEwen, L. R. Ong, and M. Zihayat, "A big data analytics framework for detecting user-level depression from social networks," *International Journal of Information Management*, vol. 54, p. 102141, 2020.

[15] J. Jia, "Mental health computing via harvesting social media data," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 7 2018, pp. 5677–5681.

[16] "Death by cause and age group and gender," Open Data Platform Singapore, https://data.gov.sg/dataset/death-by-cause-and-age-group, Last accessed on March 15, 2021.

[17] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 05 2003.

[18] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 7821–6, 07 2002.

[19] R. Rehurek and P. Sojka, "Gensim–python framework for vector space modelling," *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, vol. 3, no. 2, 2011.

[20] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text." in *ICWSM*, E. Adar, P. Resnick, M. D. Choudhury, B. Hogan, and A. H. Oh, Eds. The AAAI Press, 2014.

[21] "Empathy map," Dave Gray, July 2017, https://gamestorming.com/empathy-mapping/, Last accessed on June 15, 2021.

[22] "Hospital admission rate by sex and age, annual," Statistics Singapore, https://www.singstat.gov.sg/find-data/search-by-theme/society/health/latest-data, Last accessed on March 15, 2021.