UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Web-scraping the Expression of Loneliness during COVID-19

Permalink

https://escholarship.org/uc/item/53m546pq

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN

1069-7977

Authors

Jung, Yoonwon
Lee, Yoon Kyung
Hahn, Sowon

Publication Date

2021

Peer reviewed

Web-scraping the Expression of Loneliness during COVID-19

Yoonwon Jung (ywjung@snu.ac.kr)
Department of Psychology, Seoul National University
1, Gwanak-ro, Gwanak-gu, Seoul, 08826 Republic of Korea

Yoon Kyung Lee (yoonlee78@snu.ac.kr)
Department of Psychology, Seoul National University
1, Gwanak-ro, Gwanak-gu, Seoul, 08826 Republic of Korea

Sowon Hahn (swhahn@snu.ac.kr)
Department of Psychology, Seoul National University
1, Gwanak-ro, Gwanak-gu, Seoul, 08826 Republic of Korea

Abstract

We investigated the subjective experience of loneliness during COVID-19 by analyzing social media postings from March 2020 to January 2021. We collected text data from loneliness-related subgroups of Reddit and sampled 12787 posts that were written in ten consecutive days from each month. The results suggest that when individuals express their loneliness, they show an internal focus of attention on their emotions, desires, and cognitive appraisals rather than an external focus of attention on situations or other people. Linguistic markers of emotions expressed by lonely individuals included depression, anxiety, anger, hate, helplessness, and sadness. Also, loneliness-related topics were generally about their internal states pertinent to various social relationships, interpersonal interaction deficits, and their own lives in broad time perspectives. COVID-19 related loneliness was associated with negative appraisal of one's situation and reaching out for new relationships online.

Keywords: Loneliness; COVID-19; Natural Language Processing (NLP); N-gram Frequency; Correlated Topic Model (CTM)

Introduction

Loneliness is a negative state subjectively experienced when there is a discrepancy between the expected and actual level of social connection (Cacioppo et al., 2015; Peplau & Perlman, 1982). Studies about loneliness have suggested three dimensions of loneliness based on the relationship with others: intimate loneliness in family, a romantic partner, and close friends, relational loneliness in casual circles whom one sees regularly, and collective loneliness in social roles that rise from institutions or group membership (Cacioppo et al., 2015; Dunbar, 2014; Goodman & Epstein, 2020).

Loneliness has been one of the major reasons for both mental and physical health problems. Loneliness is negatively associated with subjective well-being, life satisfaction, and vitality (Elphinstone, 2018). Depression and anxiety are highly associated with high levels of loneliness (Hawkley & Cacioppo, 2010). Loneliness can also elevate the risk of stroke, coronary heart disease, and high blood pressure (Valtorta et al., 2016). It also increases the mortality risk by 32 percent regardless of nationality and gender (Holt-Lunstad et al., 2015).

Loneliness has been a worldwide health problem even before the outbreak of coronavirus. It affected a third of people in industrialized countries (Cacioppo & Cacioppo, 2018). Now in the era of the COVID-19 pandemic, long-term social isolation and reduced physical activities have further increased the risk of loneliness (Pinquart & Sorensen, 2001; Tanskanen & Anttila, 2016). Recent study showed that people were increasingly in a state of social isolation and felt lonelier during the period of COVID-19 (Kilgore et al., 2020). People in quarantine or emergency workers increasingly experienced social isolation and loneliness, which in turn increased other mental health problems such as anxiety and depression (Banerjee & Rai, 2020).

With the increased use of social media platforms, web-scraping is emerging as a novel approach to measure the mental state and behaviors of people. Web-scraping is a type of crowdsourcing which collects public data from existing online sources (Guntuku et al., 2017). It has the advantage of collecting larger samples in a faster way and enables the collection of naturalistic data with high ecological validity (Guntuku et al., 2017; Priniski & Holyoak., 2020).

Recently, some researchers have started to use web-scraping to investigate expressed loneliness and related behaviors. Before the outbreak of COVID-19, Guntuku et al. (2019) studied the characteristics of loneliness-related posts compared to a control group of matched age, gender, and period of posting. Then, Koh and Liew (2020) examined how expressions of loneliness have changed across time and region during the two months of the pandemic. However, both research collected twitter posts explicitly using "lonely", "loneliness", or "alone", and did not contain the data written during the prolonged period of COVID-19. More research is needed on how lonely people express various aspects of experience without directly mentioning the word "lonely" and how prolonged COVID-19 is influencing loneliness.

We investigated different aspects of loneliness experience during 10 months of COVID-19 situations by web-scraping 12787 posts from loneliness-related Reddit communities. We planned to capture the characteristics of lonely individual's language used when talking about their experiences online. It has been reported that depressed individuals show more self-focused attention (Ingram & Smith, 1984; Guntuku et al., 2017), and lonely individuals refer more to themselves and talk more of their feelings and needs compared to the control

group (Guntuku et al., 2019). Guntuku and other researchers (2019) also found that lonely Twitter users demonstrated linguistic markers of depression, anxiety, and anger. Considering the fact that the state of loneliness is associated with depression and that such association is also presented in languages that lonely people use, we speculated that 1) lonely individuals' language will show the internal focus of attention by focusing more on their own thoughts, feelings, and desires. We also predicted that 2) lonely individuals will show topics related to depression, anxiety, and anger.

Also, we expected that language used in such posts will suggest linguistic markers that reveal the nature of the subjectively experienced loneliness. We tried to explore 3) the characteristics of COVID-19-related loneliness and other various contexts in which lonely people experience difficulties, and 4) how the various experiences of loneliness are correlated with each other

Method

We selected Reddit as a social media platform for web-scraping among various options. Reddit is a community-based social media platform where users can post, comment, and vote. It is chosen for the ease of access to a large amount of user-generated content, longer texts compared to other platforms like Twitter or Quora, and the existence of subreddits specifically related to loneliness.

We searched Reddit posts from March 27th of 2020 to January 5th of 2021. We specified a total of 10 time periods, each time period being the consecutive 10 days. Specific time periods are listed in Table1.

Materials

Reddit Posts We collected posts from loneliness-related subreddits; namely, "lonely" and "loneliness" subreddits. Subreddits are communities that any user can freely create and join. Such communities are managed by volunteers who perform tasks such as deleting inappropriate content and banning harmful users. "Lonely" subreddit has 188k members and "loneliness" subreddit has 8.1k members.

Reddit posts consist of title and body. We collected both title and body texts from each post. All posts in the selected subreddits were in English.

Pushshift Application Programming Interface (API) Pushshift API enables access to the contents of Reddit posts and comments via Pushshift dataset, which stores the copy of Reddit objects at the time they are created in Reddit (Baumgartner et al., 2020). Pushshift API not only enables scraping posts in specific subreddits but also makes extracting posts between two date points possible. We collected all posts within one of the two subreddits using and created within the assigned date range.

Procedure

We conducted all procedures in the python3 environment.

Data Collection We collected posts using the Pushshift API (Baumgartner et al., 2020)[1]. The data could be accessed directly via 'https://api.pushshift.io/' endpoint, and the URL could be tailored to the specific needs of the collection. Total 13462 posts were initially collected across 10 time periods, with the number of posts from the "lonely" and "loneliness" subreddit being 11807 and 359 respectively.

N-gram Frequency We computed overall word frequency and created unigram and bigram frequency graphs to see the main trend of topics discussed among lonely individuals. Unigram is an n-gram of size 1, and bigram is an n-gram of size 2, where n-gram is a sequence of continuous n words in a given sentence. For example, in "I want to meet new friends", "friends" is a unigram, and "new friends" is a bigram. Bigrams can capture the more specific context of how words are used together to convey the intentions of writers.

In order to create unigrams and bigrams, we split the texts of each post from body to title into multiple sentences using nltk library. Then, we split each sentence into words of lowercase form, with special tokens replacing user names and emojis: <another_user> for usernames, <pos_emoji> for positive emojis, <neg_emoji> for negative emojis, and <neutral_emoji> for neutral emojis. We replaced URLs to the web page title and removed other subreddits' names.

Then, we lemmatized each word and normalized negative adverbs. Lemmatization normalizes words by converting words into the base dictionary form based on POS information. For negative adverbs, we transformed a pair of a negative adverb and the consecutive token into one token in a form that "not_" is attached before the second token. For example, we normalized "can't believe" to "not_believe".

After undergoing such procedures, we calculated unigrams and bigrams of tokens by each sentence using 'Counter' library. When either unigram or a word from bigram contained 'stopwords', they were excluded from analysis. Stop words are words that occur in high frequency but contain no significant information regarding the purpose of analysis. We created a stop words list based on the nltk stop words list. 30 most frequent unigrams and bigrams are counted to capture the general view of topics being discussed. We used 'Counter' and 'matplotlib' library for the visualization.

Topic Modeling Topic modeling is a type of unsupervised machine learning algorithm that extracts latent topics from large data. Among various topic modeling methods, we used Correlated Topic Model (CTM). CTM overcomes the limitation of widely used topic modeling algorithms such as Latent Dirichlet Allocation (LDA) and other LDA-based models. Such models fail to capture correlations between topics because of Dirichlet distribution's independence assumption (Blei & Lafferty, 2007). However, it is highly

likely that certain topics are likely to appear with the appearance of other topics, and correlations among topics in our corpus are more likely considering it being written by relatively homogenous people. Like LDA, CTM extracts the pre-defined number of latent topics from the probabilistic distribution of words. However, CTM uses logistic normal distribution in substitute of Dirichlet distribution to model the proportions of implicit topics in a document (Blei & Lafferty, 2007). In our data, we defined documents as pre-processed texts from each post.

The procedure of text pre-processing for topic modelling was mostly similar to that used when creating N-grams except for two differences. First, we combined the pre-processed form of title and body texts to form each document, instead of separating the sentences into multiple documents. Second, we removed stopwords before lemmatizing. Lists of stopwords and emojis were the same as the one used when computing N-gram frequency.

The final form of pre-processed tokens is used for building vocabs. With vocabs, which are vectorized tokens to be used as inputs for CTM, we removed vocabs that are used in less than 10 documents and conducted experiments on changing 4 hyperparameters of CTM. First, we adjusted the removal of words that appeared most frequently across all topics from 0 to top 5, with topic number set to a few constant values and all other parameters to default. Based on inter-topic discriminability, we decided to remove 5 top words. The list of removed words is: 'feel', 'friend', 'want', 'people', and 'like'. Then, we conducted 100 experiments to adjust the other three parameters: number of topics from 5 to 29, eta from 0.01 to 0.03, term weights of either uniform term weights or inverse document term weights. For all models, we conducted 300 iterations and compared c_v coherence scores to maximize the coherence among words comprising topics. Among 100 experiments, we chose 5 models with the highest coherence scores across all experiments and decided the final model based on inter-topic discriminability and human interpretability.

With the final model, we classified documents to each topic based on the highest distribution among topics in the document. Then, we measured correlation among topics using Pearson's r and visualized the correlations within the 10 percent of strongest correlation intensity. The cutoff point was $|r|=0.11016306$. We used the 'Tomotopy' v0.10.0[23] for vectorizing corpus, defining the optimal number of topics, training the final model, classifying documents, and computing correlations. We used pyvis and matplotlib library for visualizing the correlations.

Result

After removing user-deleted posts, we used 12787 posts in total for analysis, with the number of posts from "lonely" and "loneliness" subreddit being 12433 and 354 respectively.

Demographic information Although Reddit does not provide user information per subreddit, demographic information of Reddit users is available. According to Marketing Charts (2019), the usage among the male population in the U.S. (15%) was almost twice the usage of the female population (8%), and use among White (12%) and Hispanic (14%) population outnumbered the use among African-American (4%) population in the United States. The age group with the highest Reddit usage was 25-29 (23%).

N-gram Frequency Figure 1 shows 30 most frequent unigrams in data. 10 Unigrams that appeared most frequently included 'feel', 'friend', 'like', 'want', 'lonely' 'people', 'talk', 'someone', 'make', and 'life'. 4 words ('friend', 'people', 'person', 'everyone') showed lonely people's external focus, and more words like 'feel', 'like', 'want', 'need', 'life', 'lonely', 'loneliness', 'alone', and 'love' showed internal focus. We saw 'anyone' and 'someone' as words that can be both used when the writer is internally focusing and externally focusing.

Figure 2 shows 30 most frequent bigrams in data. 'feel like', 'anyone else', 'feel_lonely', 'anyone want', 'best friend', 'make_friend', 'need someone', 'want someone', 'new year', and 'year old' were 10 most prominent bigrams. Bigrams also showed more internal focus than external focus: 'feel lonely', 'feeling lonely', 'feel alone', 'feel like', 'felt like', 'make feel', 'mental health' reflected the writers' emotions or mental state, and 'want someone', 'want friend', 'want talk', 'need someone', 'someone talk', 'talk people', 'not_have anyone' showed the need for connection. Among bigrams showing external focus, 'anyone want', 'anyone else', and 'new people' reflected lonely individuals' attempts to reach out to new people online for help. Also, 'best friend', 'close friend', 'new friend', and 'good friend' reflected friendship. 'Everyone else' can reflect both the inference on others and on one's situation. 'High school', 'video game', and 'social media' showed the characteristics of the community.

Table 1: Summary of data used in study.

| Time Period	Number of posts	Average words per post
March 27th to April 5th	1303	136
April 26th to May 5th	1298	144
May 27th to June 5th	1162	133
June 26th to July 5th	1257	140
July 27th to August 5th	1407	135
August 27th to September 5th	1210	137
September 26th to October 5th	1322	132
October 27th to November 5th	1246	116
November 26th to December 5th	1140	111
December 27th to January 5th	1442	116
	12787	130
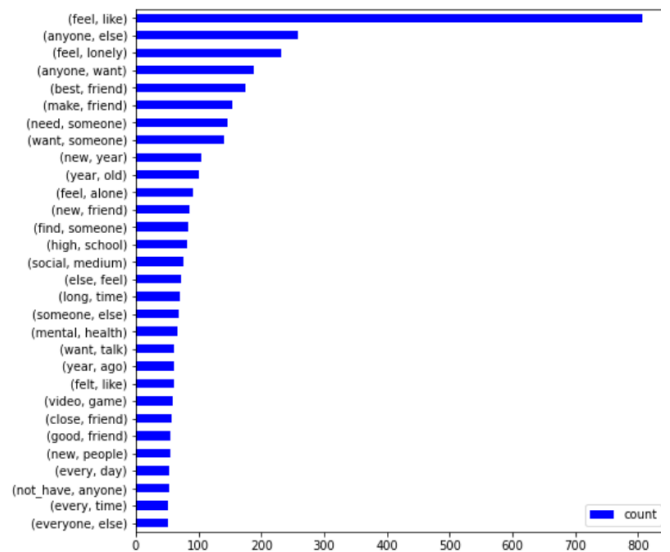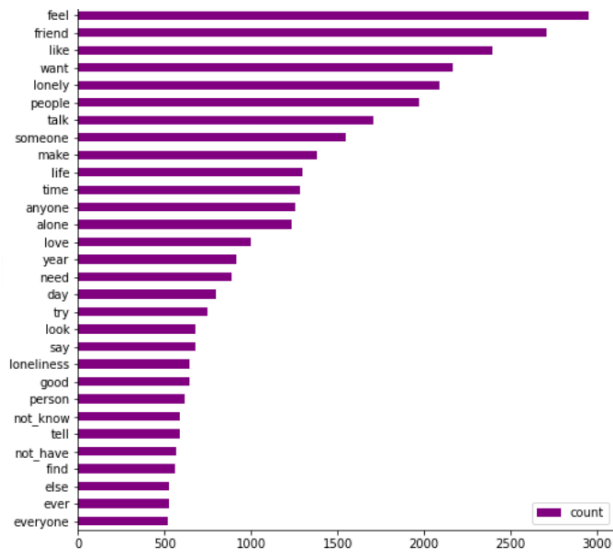
Figure 1: Unigrams of frequent words used by lonely people.

Figure 2: Bigrams of frequent words used by lonely people.

Table 2: Topics associated with loneliness.

Topic	Topic label	Top 8 keywords	Document percentage
1	Anger and hate	Girl, everything, f**k, away, hate, many, another, probably	8.10
2	Being rejected and hurt	Day, tell, not_want, hurt, let, little, stop, know	6.16
3	Feelings of loneliness	Lonely, loneliness, family, leave, not_friend, real, nice, almost	10.26
4	Feelings and thoughts during nighttime	Come, back, night, online, past, college, job, hour	4.7
5	Helplessness and sadness	Try, nothing, same, thing, nobody, good, self, cry	5.89
6	Negative experience from lacking social interaction	Ever, hard, anymore, conversation, read, life, bit, idk	4.83
7	Feeling sad due to relationship issues	Person, think, meet, sad, close, end, share, might	4.7
8	Feeling alone and longing for social connection	Alone, wish, live, move, without, today, spend, together	6.59
9	Difficulties staying connected with friends	Say, long, keep, thing, see, best, message, friendship	4.22
10	Negative appraisal on loss	Bad, way, lot, lose, maybe, time, kind, stuff	3.17
11	Desire for social connection during quarantine	Someone, love, care, post, not_feel, miss, hey, quarantine	7.17
12	One's past mental state	Well, month, few, last, point, ago, feeling, mind	2.62
13	Lack of social interaction and hoping to talk to other people	Work, since, hope, week, happen, not_talk, problem, realize	3.04
14	Past and present school life	School, use, call, home, become, hang, two, high	2.99
15	Reaching out for new relationships online	Find, look, help, give, others, great, interest, music,	2.93
16	Difficulties from experiencing anxiety	Talk, start, seem, play, game, different, anxiety, sleep	4.29
17	Looking for someone to talk with online	Time, need, felt, chat, group, pretty, listen, cause	4.0
18	Negative view on one's situation and desire to make friends	Make, anyone, good, sh*t, watch, please, situation, show	3.3
19	Finding positivity in life	Life, still, much, reason, not_care, fun, end, vent	2.59
20	"Everyone happy but me" in relationships.	Everyone, relationship, happy, right, else, parent, place, enough	3.05
21	Thoughts on how the future social interaction would be like	Year, new, social, old, around, next, write, man	2.57
22	Trying to find a new romantic relationship	Something, guy, ask, take, date, guess, sorry, girlfriend	2.92

Topic Modeling The final model used 4119 vocabs for analysis and has 22 topics with eta=0.03, uniform term weight, and coherence score of 0.420. We named the label of topics by extracting the top 8 keywords for each topic and interpreting the combination of keywords in the context of loneliness. We also read original texts from documents that are classified to each topic for labeling accuracy.

Among documents classified into 22 topics, those belonging to 'feelings of loneliness' showed the highest percentage overall of 10.26%, followed by 'anger and hate', and 'wanting for social connection during quarantine' of 8.10% and 7.17% each. The third and fourth frequent topics were 'feeling alone and longing for social connection', and 'being rejected and hurt' of 6.59% and 6.16% each. The percentages of documents belonging to the other topics were between 2% and 6%. The full topic list with keywords and the percentage of classified documents is in table2.

Among all topics, 1 topic (topic11) was associated with COVID-19 related situational changes such as social distancing and stay-at-home policy, while the other topics were not explicitly related to the pandemic. Also, 16 topics showed the writer's internal focus. 9 topics (topic1, topic2, topic3, topic4, topic5, topic7, topic8, topic12, topic16) reflected emotions, 5 topics (topic10, topic18, topic19, topic20, topic21) reflected the cognitive appraisal on one's situation, and 4 topics (topic8, topic11, topic13, topic18) represented desires. Topic 8 included both emotion and desire, and topic 18 reflected both the cognitive appraisal and desire of the writers. Among the 6 remaining topics, 3 topics (topic15, topic17, topic22) showed the behavior of the writers reaching out to other people, and 3 topics (topic6, topic9, topic14) reflected the writer's situation.
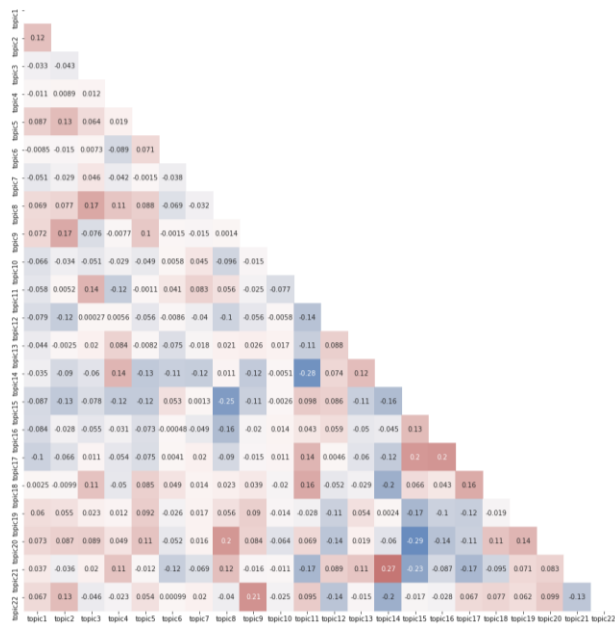
correlation.

The overall trend of correlation among topics was having 236 positive and 248 negative correlations. 'Reaching out for new relationships online' and '"Everyone is happy but me" in relationships' were r=-0.29 (topic15 and topic20), the strongest correlation in terms of absolute value. The second and third strongest correlation were: 'desire for social connection during quarantine' and 'past and present school life' (topic11 and topic14, r=-0.28), and 'past and present school life' and 'Thoughts on how future social interactions would be like' (topic14 and topic21, r=0.2). Other major correlations included: 'feeling alone and longing for social connection' and 'reaching out for new relationship online' (topic8 and topic15, r=-0.25), 'reaching out for new relationship online' and 'thoughts on how future social interactions would be like' (topic15 and topic21, r=-0.23), and 'difficulties staying connected with friends' and 'trying to find a new romantic relationship' (topic9 and topic22, r=0.22). Correlations among all topics are in figure3.

Among the top 10% of correlations (figure 4), 'past and present school life' (topic14) showed the highest number of associations with other topics, having 2 positive and 8 negative correlations. 'Reaching out for new relationship online' (topic15) and 'desire for social connection during quarantine' (topic11) followed as showing the second-highest number of correlations, with topic15 having 1 positive and 6 negative correlations and topic11 having 3 positive and 4 negative correlations. Other topics with high number of correlations included 'thoughts on how future social interactions would be like' (topic21), having 3 positive and 3 negative correlations, and 'being rejected and hurt' (topic2) and 'feeling alone and longing for social connection' (topic8) both having 3 positive and 2 negative correlations.

Figure 3: Correlation coefficients among all topics in Pearson's r. Blue and red show the negative and positive correlation each. High color saturation represents a stronger
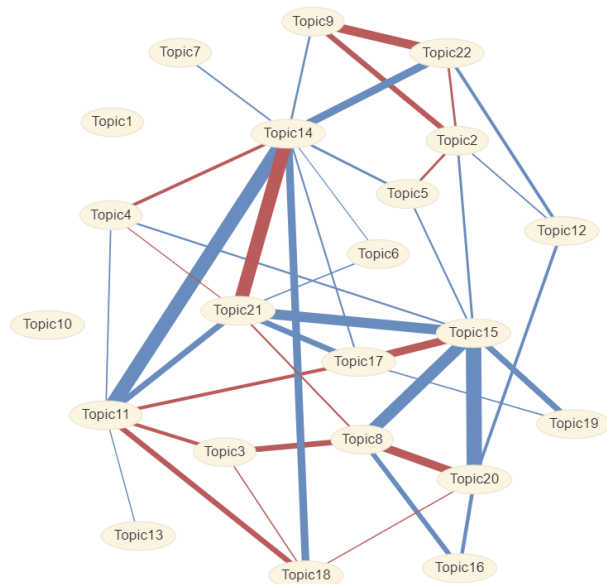

Figure 4: Visualization of top10% correlations among topics. Blue and red nodes represent the negative and positive correlation in Pearson's r. Higher thickness of nodes reflects

a stronger correlation.

A topic containing a specific COVID-19 topic word 'quarantine', 'desire for social connection during quarantine' (topic11), showed 3 major positive correlations with other topics: 'feelings of loneliness' (topic11 and topic3, r=0.14), 'negative view on one's situation and desire to make friends' (topic11 and topic18, r=0.16), and 'looking for someone to talk with online' (topic 11 and topic17, r=0.14).

Based on figure4, we grouped topics that are positively correlated with each other to infer themes emerging from the relationships between topics. The first cluster consists of topic2, topic5, topic9, topic22. Lonely people who talked about 'being rejected and hurt' were likely to talk about 'helplessness and sadness' (topic2 and topic5, r=0.13), 'wanting for social connection during quarantine' (topic2 and topic9, r=0.17), and 'trying to find a new romantic relationship' (topic2 and topic22, r=0.13). 'Wanting for social connection during quarantine' was also positively associated with 'trying to find a new romantic relationship' (topic9 and topic22, r=0.21).

The second cluster consists of topic3, topic8, topic11, topic18, and topic20. Lonely individuals who expressed 'feelings of loneliness' online were likely to express 'feeling alone and longing for social connection' (topic3 and topic8, r=0.17), 'desire for social connection during quarantine' (topic3 and topic11, r=0.14), and show 'negative view toward one's situation and desire for social connection' (topic3 and topic 18, r=0.11). Lonely writers who felt 'everyone being happy but me in relationships' were likely to reveal 'feeling alone and longing for social connection' (topic20 and topic8, r=0.2), and 'negative view toward one's situation and desire for social connection' (topic20 and topic18, r=0.11). Showing 'desire for social connection during quarantine' was also positively associated with 'negative view toward one's situation and desire for social connection' (topic11 and topic 18, r=0.16).

The third cluster consists of topic4, topic14, topic21. Writing about 'feelings and thoughts during nighttime' were associated with 'past and present school life' (topic4 and topic14, r=0.14), and 'thoughts on how future social interactions would be like' were positively correlated with both topics (topic21 and topic4, r=0.11; topic21 and topic14, r=0.27). The three clusters were generally uncorrelated with each other, except for the positive correlation between 'feeling alone and longing for social connection' and 'thoughts on how the social interaction would be like' (topic8 and topic21, r=0.12).

Negative correlations were found mainly in relation to 'past and present school life' (topic14) and 'reaching out for new relationships online' (topic15), each having 8 and 6 negative correlations with other topics. Also, there were topics that showed negative correlations with the topics that were positively correlated with each other. 'Thoughts on how future social interactions would be like' were negatively correlated with 'desire for social connection during quarantine' (topic21 and topic11, r=-0.17), 'looking for someone to talk with online' (topic21 and topic17, r=-0.17),

and 'helplessness and sadness' (topic21 and topic15, r=-0.23). 'One's past mental state' was negatively associated with 'being rejected and hurt' (topic12 and topic2, r=-0.12), and 'trying to find a new romantic relationship' (topic12 and topic22, r=-0.14). 'Difficulties from experiencing anxiety' showed negative correlation with 'feeling alone and longing for social connection' (topic16 and topic8, r=-0.16), '"Everyone happy but me" in relationships' (topic16 and topic20, r=-0.14).

Discussion

This study used a social media platform to see how people express loneliness in COVID-19 situations. We analyzed posts written in "lonely" and "loneliness" subreddits using natural language processing techniques including N-gram frequency analysis and Correlated Topic Model.

We found that lonely individuals show internal attentional focus when sharing loneliness-related experiences. The writers expressed desires, emotions, and their subjective cognitive appraisal more than describing other people or their situations. Also, the frequent emotions expressed were loneliness, anger, hate, helplessness, sadness, and anxiety. It can be concluded that our first and second hypothesis was supported, replicating the previous studies.

Among topics that lonely people talked about online, we were able to identify and label three distinct clusters based on the major top 10% positive correlations among topics. We identified the first cluster as feelings and desires related to one's past or new relationships, the second cluster of topics as feelings, desires, and cognition when lacking social interaction, and the third cluster of topics as feelings and thoughts when reflecting on one's past and imagining one's future. Such clusters of topics suggest that lonely people during pandemic generally talked about their feelings, desires, and appraisals related to their relationship and lacking social interaction, and thought about their own lives in broad time perspectives.

We observed how people have been experiencing COVID-19 related loneliness. Desire for social connection during quarantine entailed loneliness, negative appraisal of one's situation, and reaching out for new relationships online. However, topic11 was the only one that contained topic words directly referring to COVID-19 related words. One possible explanation is that a strong internal focus of attention may have led lonely individuals to write less about the COVID-19 situation itself but focus more on their internal states. Or another possibility is that the lonely Reddit communities have a significant percentage of individuals suffering from chronic loneliness, who had been lonely even before the outbreak of COVID-19. For example, topic21 was negatively associated with topic11 and was positively associated with topic14 and topic4. It suggests that picturing what one's future social life would be like did not relate to the changes caused by the pandemic but in the context of one's negative past experiences and emotions felt during the nighttime. This may reflect the existence of lonely users who go through negative experiences regardless of COVID-19

situations in the communities.

Finally, it is worth noting that lonely individuals showed desires and behaviors to reach out for new relationships via an online platform. This suggests that social media not only functions as a place to express loneliness and receiving consolation and empathy from other lonely people but also as a platform that lets individuals try new relationships. However, since the topic of 'reaching out for new relationship online' showed one of the most frequent negative correlations with other topics and the only positive correlation existing was with 'looking for someone to talk with online', the group of people who try to connect with new others online may be separable from other users who express difficulties regarding their loneliness experiences.

Our results driven from unstructured, free-written texts will help understand the nature of loneliness, and the characteristics of loneliness in COVID-19 era. Future research could compare the language use of lonely people with the control group. When using crowdsourcing based on online communities, the differences in topic distributions between lonely groups and groups interested in other plain topics could be investigated. Lonely groups in English-speaking culture could also be compared to other non-western cultures for capturing the difference in loneliness between cultures. Lastly, it should be noted that Reddit users are skewed in gender, race, and age. Our data could reflect the characteristics that are specific to the platform.

References

Banerjee, D., & Rai, M. (2020). Social isolation in Covid-19: The impact of loneliness. *International Journal of Social Psychiatry*, *66*(6), 525–527.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 830-839.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of *science*. *The Annals of Applied Statistics*, *1*(1), 17-35.

Cacioppo, J. T., & Cacioppo, S. (2018). The growing problem of loneliness. *The Lancet*, *391*(10119), 426. https://doi.org/10.1016/S0140-6736(18)30142-9

Cacioppo, S., Grippo, A. J., London, S., Goossens, L., & Cacioppo, J. T. (2015). Loneliness: Clinical import and interventions. *Perspectives on Psychological Science*, *10*(2), 238–249.

Dunbar, R. I. M. (2014). The social brain: Psychological underpinnings and implications for the structure of organizations. *Current Directions in Psychological Science*, *23*(2), 109-114.

Elphinstone, B. (2018). Identification of a suitable short-form of the UCLA-Loneliness scale. *Australian Psychologist*, *53*(2), 107-115.

Goodman, L. A., & Epstein, D. (2020). Loneliness and the COVID-19 pandemic: Implications for intimate partner violence survivors. *Journal of Family Violence*, 1-8.

Guntuku, S. C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., Polsky, D., Volpp, K. G., & Merchant, R. (2019). Studying expressions of loneliness in individuals using twitter: An observational study. *BMJ Open*, *9*(11), Article e030355.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43-49.

Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, *40*(2), 218-227.

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, *10*(2), 227-237.

Ingram, R. E., & Smith, T. W. (1984). Depression and internal versus external focus of attention. *Cognitive Therapy and Research*, *8*(2), 139-151.

Kilgore, W. D. S., Cloonan, S. A., Taylor, E. C., & Dailey, N. S. (2020). Loneliness: A signature mental health concern in the era of COVID-19. *Psychiatry Research, 290*, 1–3.

Koh, J. X., & Liew, T. M. (2020). How loneliness is talked about in social media during COVID-19 pandemic: Text mining of 4,492 Twitter feeds. *Journal of Psychiatric Research*.

Marketing Charts. (2019, April 24). *Social networking platforms' user demographics update 2019*. https://www.marketingcharts.com/digital/social-media-108184

Peplau, L. A., & Perlman, D. (1982). Perspectives on loneliness. In L. A. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy*. Wiley.

Pinquart, M., & Sorensen, S. (2001). Influences on loneliness in older adults: A meta-analysis. *Basic and Applied Social Psychology*, *23*(4), 245-266.

Priniski, J. A., & Holyoak, K. J. (2020). Crowdsourcing to analyze belief systems underlying social issues. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Tanskanen, J., & Anttila, T. (2016). A prospective study of social isolation, loneliness, and mortality in Finland. *American Journal of Public Health*, *106*(11), 2042-2048.

Valtorta, N. K., Kanaan, M., Gilbody, S., Ronzi, S., & Hanratty, B. (2016). Loneliness and social isolation as risk factors for coronary heart disease and stroke: Systematic review and meta-analysis of longitudinal observational studies. *Heart*, *102*(13), 1009-1016.